PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/48 A3

(11) International Publication Number: WO 99/50385

(43) International Publication Date: 7 October 1999 (07.10.99)

(21) International Application Number: PCT/US99/06892

(22) International Filing Date: 30 March 1999 (30.03.99)

(30) Priority Data:
60/079,928 30 March 1998 (30.03.98) US

(71) Applicant (*for all designated States except US*): PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BELLAICHE, Yohanns [FR/FR]; 46, rue d'Ulm, F-75005 Paris (FR). THE, Siu Inge [NL/US]; 88 Worcester Street #5, Boston, MA 02118 (US). PERRIMON, Norbert [FR/US]; 6 Frost Street, Arlington, MA 02174 (US).

(74) Agents: CELLA, Charles, H. et al.; Foley, Hoag & Eliot, LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

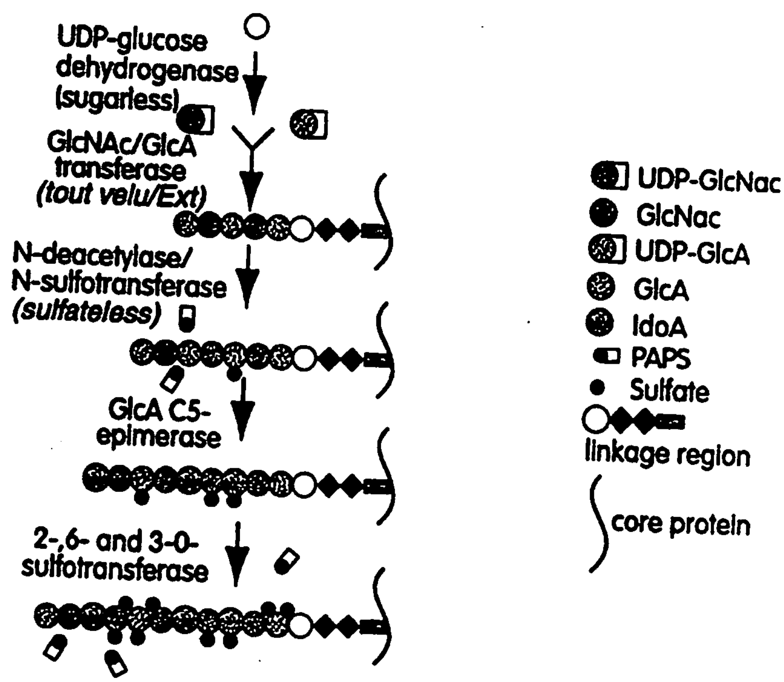
Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 18 November 1999 (18.11.99)

(54) Title: REGULATION OF GLYCOSAMINOGLYCAN SYNTHESIS, METHODS AND REAGENTS RELATED THERETO

(57) Abstract

The present invention concerns the discovery of a new family of *hedgehog* interacting proteins, referred to herein as "*hedgegehog* interacting proteins" or "*Exts*", which are demonstrated to bind to *hedgehog* polypeptides with high affinity. As described herein, the *Ext* proteins are required for and as such regulate hedgehog diffusion.



38

USSN # 09/847,102
Docket # 220002062900

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/06892

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/48

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 541 095 A (HIRSCHBERG CARLOS B ET AL) 30 July 1996 (1996-07-30) the whole document ---	1,15
A	WO 94 09037 A (UNIV NEW YORK) 28 April 1994 (1994-04-28) the whole document ---	1,15
P,A	WO 98 30576 A (BEACHY PHILIP A ;PORTER JEFFREY A (US); UNIV JOHNS HOPKINS MED (US)) 16 July 1998 (1998-07-16) examples ---	1,15
P,A	WO 98 52553 A (UNIV JOHNS HOPKINS) 26 November 1998 (1998-11-26) examples ------	1,15

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 September 1999	06/10/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Moreno, C

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/06892

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5541095	A	30-07-1996	AU	4529593 A	04-01-1994
			WO	9325659 A	23-12-1993
WO 9409037	A	28-04-1994	AU	5443394 A	09-05-1994
			CA	2147167 A	28-04-1994
			EP	0677063 A	18-10-1995
			JP	8502487 T	19-03-1996
			NZ	257713 A	28-10-1996
WO 9830576	A	16-07-1998	AU	4800697 A	03-08-1998
WO 9852553	A	26-11-1998	AU	7686698 A	11-12-1998

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N A2

(11) International Publication Number: WO 99/50385

(43) International Publication Date: 7 October 1999 (07.10.99)

(21) International Application Number: PCT/US99/06892

(22) International Filing Date: 30 March 1999 (30.03.99)

(30) Priority Data:
60/079,928 30 March 1998 (30.03.98) US

(71) Applicant (*for all designated States except US*): PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): BELLAICHE, Yohanns [FR/FR]; 46, rue d'Ulm, F-75005 Paris (FR). THE, Siu Inge [NL/US]; 88 Worcester Street #5, Boston, MA 02118 (US). PERRIMON, Norbert [FR/US]; 6 Frost Street, Arlington, MA 02174 (US).

(74) Agents: CELLA, Charles, H. et al.; Foley, Hoag & Eliot, LLP, One Post Office Square, Boston, MA 02109 (US).

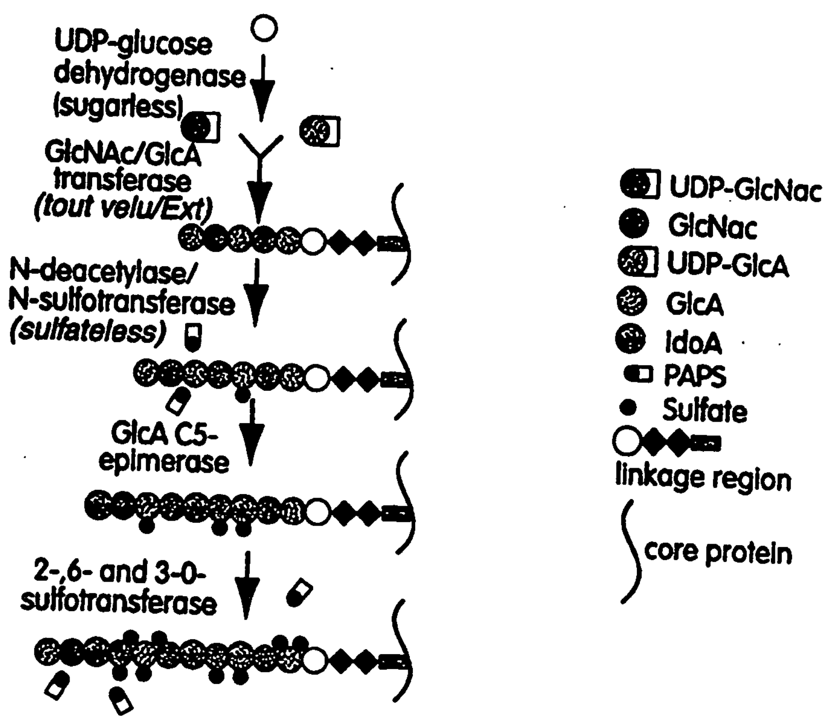
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: REGULATION OF GLYCOSAMINOGLYCAN SYNTHESIS, METHODS AND REAGENTS RELATED THERETO

(57) Abstract

The present invention concerns the discovery of a new family of *hedgehog* interacting proteins, referred to herein as "*hedgegehog* interacting proteins" or "*Exts*", which are demonstrated to bind to *hedgehog* polypeptides with high affinity. As described herein, the *Ext* proteins are required for and as such regulate hedgehog diffusion.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

Regulation of Glycosaminoglycan Synthesis, Methods and Reagents Related Thereto

Background of the Invention

Growth factors and cytokines are important signaling molecules that control many physiological processes such as cell growth, division, differentiation, and migration. These molecules constitute some of the most important regulators of gene expression in eukaryotic systems. Much of the work to date focuses on the downstream effects of these molecules on cells, such as the initiation of signal transduction through the binding of the growth factor to its transmembrane receptor or the different second messenger pathways employed by certain signaling molecules. However, little attention has been paid to how the growth factors may be regulated through interactions with the ECM, from the time of their release until they form an active complex with their receptors at the cell surface.

Understanding of ECM-growth factor interaction is especially critical in light of the fact that an increasing number of growth factors that have been found to form strong complexes with the ECM, specifically glycosaminoglycan-containing proteoglycans. For example, members of the fibroblast growth factor (FGF) family require the presence of glycosaminoglycan (GAG) chains to act as a cofactor for binding to, and activation of, FGF-receptors on the cell surface. Platelet factor 4 binds heparin and is complexed to the preoteoglycan serglycin upon secretion from platelet -alpha granules.

Membrane associated carbohydrate is exclusively in the form of oliogsaccharides covalently attached to proteins forming glycoproteins, and to a lesser extent covalently attached to lipid forming the glycolipids. Glycoproteins consist of proteins covalently linked to carbohydrate. The predominant sugars found in glycoproteins are glucose, galactose, mannose, fucose, GalNAc, GlcNAc and NANA. The distinction between proteoglycans and glycoproteins resides in the level and types of carbohydrate modification. The carbohydrate modifications found in glycoproteins are rarely complex: carbohydrates are linked to the protein component through either O-glycosidic or N-glycosidic bonds. The N-glycosidic linkage is through the amide group of asparagine. The O-glycosidic linkage is to the hydroxyl of serine, threonine or hydroxylysine. The linkage of carbohydrate to hydroxylysine is generally found only in the collagens. The linkage of carbohydrate to 5-hydroxylysine is either the single sugar galactose or the disaccharide

glucosylgalactose. In ser- and thr-type O-linked glycoproteins, the carbohydrate directly attached to the protein is GalNAc. In N-linked glycoproteins, it is GlcNAc.

The synthesis of O-linked glycoproteins occurs via the stepwise addition of nucleotide-activated sugars directly onto the polypeptide. The nucleotide-activated sugars are coupled to either UDP, GDP (as with mannose) or CMP (for instance, NANA). The attachment of sugars is catalyzed by specific glycoprotein glycosyltransferases. Evidence indicates that each specific type of carbohydrate linkage in O-linked glycoproteins is the result of a different glycosyltransferase.

As indicated earlier, the three major classes of N-linked carbohydrate modifications are high-mannose, hybrid and complex. The major distinguishing feature of the complex class is the presence of sialic acid, whereas the hybrid class contains no sialic acid.

The most abundant heteropolysaccharides in the body are the glycosaminoglycans (GAGs). These molecules are long unbranched polysaccharides containing a repeating disaccharide unit. The disaccharide units contain either of two modified sugars: N-acetylgalactosamine (GalNAc) or N-acetylglucosamine (GlcNAc), and a uronic acid such as glucuronate or iduronate. GAGs are highly negatively charged molecules, with extended conformation that imparts high viscosity to the solution. GAGs are located primarily on the surface of cells or in the extracellular matrix (ECM). Along with the high viscosity of GAGs comes low compressibility, which makes these molecules ideal for a lubricating fluid in the joints. At the same time, their rigidity provides structural integrity to cells and provides passageways between cells, allowing for cell migration. The specific GAGs of physiological significance are hyaluronic acid, dermatan sulfate, chondroitin sulfate, heparin, heparan sulfate, and keratan sulfate. Although each of these GAGs has a predominant disaccharide component, heterogeneity does exist in the sugars present in the make-up of any given class of GAG.

The distinction between proteoglycans and glycoproteins resides in the level and types of carbohydrate modification. The carbohydrate modifications found in glycoproteins are rarely complex: carbohydrates are linked to the protein component through either O-glycosidic or N-glycosidic bonds. The N-glycosidic linkage is through the amide group of asparagine. The O-glycosidic linkage is to the hydroxyl of serine, threonine or hydroxylysine. The linkage of carbohydrate to hydroxylysine is generally found only in the collagens. The linkage of carbohydrate to 5-hydroxylysine is either the single sugar galactose or the disaccharide glucosylgalactose. In ser- and thr-type O-linked

glycoproteins, the carbohydrate directly attached to the protein is GalNAc. In N-linked glycoproteins, it is GlcNAc.

The majority of GAGs in the body are linked to core proteins, forming proteoglycans (also called mucopolysaccharides). The GAGs extend perpendicularly from the core in a brush-like structure. The linkage of GAGs to the protein core involves a specific trisaccharide composed of two galactose residues and a xylulose residue (GAG-GalGalXyl-O-CH2-protein). The trisaccharide linker is coupled to the protein core through an O-glycosidic bond to a S residue in the protein. Some forms of keratan sulfates are linked to the protein core through an N-asparaginyl bond. The protein cores of proteoglycans are rich in S and T residues, which allows multiple GAG attachments.

The most important subset of the proteoglycans is derived with heparin-sulfate glycosaminoglycans (HS-GAGs), complex acidic polysaccharide polymers consisting of a disaccharide repeat unit of glucosamine and a uronic acid. Previous attempts to elucidate structure-function relationships for HS-GAGs in modulating biological activity have been thwarted because of the diverse chemical structure of the polymer. The glucosamine/ uronic acid repeat unit of HS-GAGs can be modified at five different sites within the repeat unit, including O-sulfation, N-sulfation or acetylation, and C5 epimerization of the uronic acid moiety. Together these modifications, as well as the unique conformational flexibility of the uronic acid component, create a highly heterogeneous polymer with diverse *in vivo* functions.

The carbohydrate backbone in heparin and HS consists of alternating hexuronic acid (D-glucuronic acid (GlcA) or L-iduronic acid (IdoA)) and D-glucosamine (GlcN) units. The GlcN residues in heparin are predominantly N-sulfated, whereas those in HS show a more varied N-substitution pattern with appreciable proportions of both N-sulfated and N-acetylated and a smaller amount of N-unsubstituted GlcN units. These structures are generated through the formation of a [GlcA-GlcNAc]n polymer that is subsequently modified by partial N-deacetylation/N-sulfation of GlcNAc units, C-5 epimerization of GlcA to IdoA residues, and, finally, incorporation of O-sulfate groups at various positions (Figure 5).

Large numbers of proteins in animal tissues occur immobilized in the extracellular space, at cell surfaces or in the extracellular matrix. Some are anchored through interactions with other proteins. However, current research increasingly implicates proteoglycans as scaffold structures, designed to accommodate proteins through

noncovalent binding to their glycosaminoglycan side chains. In particular, heparan sulfate proteoglycans are recognized as ubiquitous protein ligands. Binding of proteins to HS chains may serve a variety of functional purposes, from simple immobilization or protection against proteolytic degradation to distinct modulation of biological activity. Because of such interactions HS proteoglycans are critically involved in a variety of biological phenomena at various levels of complexity, including organogenesis in embryonic development, angiogenesis, regulation of blood coagulation and growth factor/cytokine action, cell adhesion, lipid metabolism, etc.

Summary of the Invention

The present invention is directed to the discovery that there exists proteoglycans having specific glycosaminoglycan chains which provide exquisite selectivity for binding to particular growth factors or cytokines, e.g., growth factor-specific GAG chains. Moreover, the specificity of the GAG chain structure, which gives rise to this binding specificity, is regulated by a biosynthetic pathway which can be selectively inhibited in order to antagonize the activity of a specific growth factor.

Based on the above understanding, one aspect of the present invention provides drug discovery assays for identifying agents which can selectively alter the activity of a growth factor, cytokine or other serum proteins (collectively referred to hereinafter as "factors" for ease of reading). In particular, the subject assay is designed to identify inhibitors of specific modifications of GAG chains, such as particular enzymes involved in, e.g. isomerization, sulfation, or acetylation which produce factor-specific binding by the resulting proteoglycan. Such inhibitors can be used to alter the specificity of a proteoglycan for binding a particular factor (or family of related factors) by inhibiting specific modifications to the sugar sequence of the proteoglycan. In one embodiment, the assay identifies agents which selectively inhibit a sulfotransferase, preferably an N-deacetylase/N-sulfotransferase, involved in synthesis of a proteoglycan that binds a selected factor - so as to decrease the ability of the proteoglycan to bind the factor.

Another aspect of the present invention relates to a method for inhibiting the activity of a factor, *in vitro* or *in vivo*, by administering a compound which selectively inhibits synthesis of the factor-specific glycosaminoglycan in an amount sufficient to reduce the responsiveness of the treated cell(s) to the factor. Such methods can be used to treat, e.g., unwanted cell proliferation or other unwanted effects of a factor.

Yet another aspect of the present invention provides a method for potentiating the activity of a selected factor, *in vitro* or *in vivo*, by administering a compound which selectively inhibits synthesis of GAG chains which bind other factors that compete with the selected factor, without substantially diminishing the synthesis of GAG chains which bind the selected factor. In such a manner, the activity of the factor can be potentiated, e.g., its mitogenic, trophic or other activity enhanced.

In one embodiment, this invention provides drug discovery assays for identifying agents which can modulate the biological function of selected factors, such as by (i) inhibiting the glycosyltransferase activity of an Ext protein, e.g. Ext1; (ii) inhibiting a GAG-modifying activity of an enzyme associated with an Ext protein, e.g. a sulfotransferase; or (iii) inhibiting the interaction of an Ext protein with other enzymes involved in GAG synthesis, e.g. sulfotransferases. For example, the subject method can be used to inhibit hedgehog-mediated signal transduction by inhibiting synthesis of hedgehog specific GAG chains, e.g., by Ext-dependent pathways. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a tissue otherwise responsive to hedgehog proteins. For example, the subject inhibitors can be used to prevent hedgehog-dependent proliferation of cells, such as hedgehog-dependent basal cell carcinoma (BCC) or other hedgehog-dependent epithelia proliferative disorders. In another example, the subject inhibitors can be used to alter the growth state of a mesodermally-derived tissue, such as cartilage. In particular, this invention provides a method for identifying agents that are effective in amielorating at least one symptom of a disorder characterized by aberrant bone morphogenesis, such as multiple exostoses, chondrosarcomas, and/or osteosarcomas. In other embodiments, the invention also helps in identifying agents useful therapeutically to inhibit hedgehog-mediated growth, maintenance and/or differentiation of a tissue involved in spermatogenesis and tissue derived from dorsal mesoderm; ectodermally-derived tissue, such as tissue derived from the epidermis, neural tube, neural crest, or head mesenchyme; endodermally-derived tissue, such as tissue derived from the primitive gut.

In another embodiment, this invention provides drug discovery assays for identifying agents which can modulate the biological function of selected factors which bind to frizzled receptors, such as Wnt proteins. For example, the assays can be used to identify inhibitors of Wnt activity by (i) inhibiting Wnt-specfic modification of glypicans by a GAG-modifying activity, e.g. inhibiting a glypican-modifying N-deacetylase/N-

sulfotransferase; (ii) inhibiting the interaction of a Wnt protein with a frizzled/glypican complex; or (iii) inhibiting formation of a a frizzled-glypican complex capable of binding to a Wnt protein. Accordingly, the subject method can be used to inhibit Wnt-mediated signal transduction by inhibiting synthesis of Wnt-specific GAG chains, e.g., present on Wnt-specific glypicans. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a tissue otherwise responsive to Wnt proteins. The identification of modulators of Wnt activity are useful in treating disease states involving the Wnt activity. Such compounds could be of use in the treatment of diseases in which activation or inactivation of the Wnt protein results in either cellular proliferation, cell death, nonproliferation, induction of cellular neoplastic transformations or metastatic tumor growth and hence could be used in the prevention and/or treatment of cancers such as bone and breast cancer for example.

In preferred embodiments, the steps of the assay are repeated for a variegated library of at least 100 different test compounds, more preferably at least 10^3, 10^4 or 10^5 different test compounds. The test compound can be, e.g., a peptide, a nucleic acid, a carbohydrate, a small organic molecule, or natural product extract (or fraction thereof).

In yet another embodiment, the present invention provides diagnostic and therapeutic assays and reagents for detecting and treating disorders involving, for example, aberrant expression (or loss thereof) of an Ext glycosyltransferase gene, abberant expression of a *dally* N-deacetylase/N-sulfotransferase gene, or abberant expression of a glypican.

The present invention further contemplates the pharmaceutical formulation of one or more agents identified in such drug screening assays.

In other embodiments, the present invention provides a molecule, preferably a small organic molecule, which is identified in the subject drug screening assays.

Yet another aspect of the present invention concerns a method for modulating one or more of growth, differentiation, or survival of a cell by treatment with a compound identify by the subject drug screening assay, e.g., by potentiating or inhibiting certain glycosyltransferase involved in factor-selective GAG synthesis. In general, whether carried out *in vivo*, *in vitro*, or *in situ*, the method comprises treating the cell with an effective amount of such an agent so as to alter, relative to the cell in the absence of treatment, at least one of (i) rate of growth, (ii) differentiation, or (iii) survival of the cell.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1: *ttv* affects Hh signaling

All discs are oriented as in a: A, anterior; P, posterior; V, ventral and D, dorsal. Blue lines indicate the A/P boundary. Dashed lines indicate the limit of the relevant mutant clones as determined by the absence of P-Galactosidase (-Gal) staining (see below). The $ttv^{l(2)00681}$ allele, which behaves as a genetic null and is a protein null (see figure 3d), was used to generate *ttv* somatic mutant clones.

a. Wild type Ptc expression, in red, is detected with the Apal3 Mouse monoclonal antibody (Capdevila, J., Pariente, F., Sampedro, J., Alonso, J.L. & Guerrero, I. *Development* Subcellular localization of the segment polarity protein patched suggests an interaction with the wingless reception complex in Drosophila embryos. 120, 987-998 (1994)). Ptc is expressed at low levels throughout the wing disc and its expression is greatly increased in the A compartment along the A/P boundary in response to Hh signaling. Note that the staining is weaker in the central region of the wing blade.

b. Wild type Ci expression, in green, is detected with rat monoclonal antibody 2AI which has a higher affinity for the full length Ci gene product than for the repressor form (Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C. & Kornberg, T.B. *Cell* Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. 89, 1043-1053 (1997)). Ci is only expressed in the A compartment and it can be used as a marker for the A compartment. The full length Ci protein is stabilized in response to Hh signaling as seen by a more intense staining. Note that the posterior limit of the domain of Ci stabilization does not correspond with the A/P boundary in late third instar wing discs due to the expression of Engrailed/Invected (En/Inv) in the A compartment. The expression of En/Inv have been shown to be dependent on high level of Hh signaling (Strigini, M. & Cohen, S.M. *Development* A

Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)).

c, d, and e. Ptc expression in red (c, e) and -Gal in green (c, d). c shows the full disc where the relevant *ttv* clone is indicated by an arrow. d and e are magnifications of the clone. The absence of -Gal staining is used as a marker for the *ttv* mutant clone. In a *ttv* clone, *ptc* expression is not induced, except in one row of cells adjacent to the A/P boundary. This staining is also reduced in comparison with wild type cells.

f, g, h, i, j and k. Ci staining in green (f, h, i, k) and -Gal staining in red (f, g, i, j). The absence of -Gal staining is used as a marker for the clone.

f, g and h. f shows the full disc where the relevant *ttv* clone is indicated by an arrow. g and h are magnifications of the clone. In homozygous *ttv* mutant cells, Ci stabilization is not induced except at the posterior edge of the clone.

i, j and k. i shows the full disc where the relevant *ttv* clone is indicated by an arrow. j and k are magnifications of the clone. k, Wild type cells anterior to a *ttv* mutant clone do not stabilize Ci, although they are in a domain competent to respond to Hh signaling (see arrowheads). The cell nonautonomous effect is directional because cells posterior to the clone respond to Hh signaling (see also h). Within the clone the distance between the domain of Ci stabilization and the A/P boundary is reduced. This indicates that the level of Hh signaling is reduced since this distance is dependent on high levels of Hh signaling (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)). Finally we interpret that the stabilization of Ci protein in the ventral cells of *ttv* clone is due to the diffusion of the Hh protein from wild type cell located more ventrally (see arrow).

Figure 2: <u>*ttv* is required for Hh diffusion</u>

Blue lines indicate A/P boundary. Dashed lines indicate the limits of the clones, unless they overlap with the A/P boundary. All discs are oriented as in Figure la. $ttv^{l(2)00681}$ and ptc^{IW} alleles were used to generate the somatic mutant clones.

In a and b, overlays of Ptc and -Gal stainings are shown in green and red, respectively are shown. In c and d, overlays of Ptc staining in green, and Hh staining in red are shown. Hh is detected with a rabbit anti-Hh antibody (Taylor, A.M., Nakano, Y., Mohler, J. & Ingham, P.W. *Mech Dev* .Contrasting distributions of patched and hedgehog proteins in the Drosophila embryo. 42, 89-96 (1993)).

a, A ptc mutant clone at the A/P boundary causes an ectopic expression of Ptc in wild type cells anterior to the clone reflecting the diffusion of the Hh protein across the ptc clone (see also (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996))).

b, Two independent *ptc, ttv* double mutant clones are shown in both the A compartment (see arrow) and the P compartment. In an A *ptc, ttv* double mutant clone, *ptc* expression is not induced in wild type cells anterior to the clone. In a P *ptc, ttv* double mutant clone, the Hh diffusion is not affected since *ptc* expression is induced anterior to the A/P border.

c, *ptc* mutant clone in the A compartment. Ptc is detected as a punctate staining, that reflects Ptc localization in vesicles. Hh is expressed in the P compartment. In a *ptc* mutant clone Hh is detected as a diffuse and as a punctate staining. When Hh reaches the wild type cells, its staining becomes more punctate and mostly coincides with punctate Ptc staining.

d. A ptc ttv double mutant clone in the A compartment. Unlike what is observed in a *ptc* mutant clone alone, Hh staining cannot be detected in *ptc ttv* mutant clones. The clone boundaries were identified by staining for P-Gal (not shown).

e, f, g. Ci staining in green (e, g) and -Gal staining in red (e, f). e shows the full disc with the relevant posterior clone indicated by an arrow. f and g are magnification of the clone. The absence of -Gal staining is used as a marker for the clone. Loss of *ttv* activity in the P compartment does not affect the diffusion of Hh in the A compartment since Ci is still stabilized in the A compartment in wild type cells adjacent to a *ttv* mutant clone induced in the P compartment.

In a wild type wing disc, Ci stabilization is not detected directly adjacent to the A/P boundary (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)). This shift (see bars |—|) in the domain of Ci stabilization depends upon high levels of Hh signaling (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)). While *ttv* clone in the P compartment does not affect the range of Hh diffusion, we noticed that the domain of Ci stabilization is now adjacent to the A/P boundary (see arrow). We interpret this observation as a reduction of the level of Hh signaling in the A compartment. This reduction in the level of Hh signaling reinforces our hypothesis of the function of Ttv

in Hh diffusion. In fact, since Hh is expressed in the P compartment, Hh signaling probably depends on the cumulative expression of Hh from several rows of cells adjacent to the A/P boundary. Since our data clearly indicate that Hh does not diffuse in *ttv* mutant cells, we expect that in a posterior *ttv* mutant clone, only the Hh protein produced by the row of cells adjacent to the A/P boundary can diffuse into the A compartment and contribute to signaling, thus explaining the shift in Ci stabilization pattern.

Figure 3: Cloning of the *ttv* locus.

a. Genomic map of the *ttv* locus. The P-element, *ttv*$^{l(2)00681}$ maps to the 51A polytene chromosome band which is deleted in *Df(2R)Trix*. It is inserted in the sixth intron of the *ttv* gene, which also includes the *lamin* C gene. Only introns of more than 1 kb are represented. Dl, G10 and B12 are three overlapping cosmids that were previously mapped to the region (Dickson, B.J., Dominguez, M., van der Straten, A. & Hafen, E. *Cell* .Control of Drosophila photoreceptor cell fates by phyllopod, a novel nuclear protein acting downstream of the Raf kinase. 80, 453-462 (1995))..

b. Sequence comparison between the Ttv and human Ext-I proteins. Black boxes indicate identical amino acids between Ext-1 and Ttv. The overall identity is 56% between Ext-1 and Ttv and 25% between Ext-2 and Ttv (not shown).

c. A Northern blot of poly-A+ RNA from 0-1.5 hour embryos probed with the *ttv* full length cDNA- A transcript of approximately 3.8 kb is detected, indicating that *ttv* is maternally expressed. *In situ* hybridization with an anti-sense *ttv* cDNA probe indicates that the *ttv* RNA is expressed ubiquitously (not shown).

d. Western blot of lysates of unfertilized wild type (WT) eggs and of eggs derived from females carrying *ttv*$^{l(2)00681}$ homozygous germ line clones (GLC) probed with an affinity purified rabbit polyclonal antibody against Ttv. The *ttv* gene encodes a 760 amino acid protein with an expected MW of 87kD. The antibody recognizes a band of approximately 80 kD which is present in wild type and absent in *ttv*$^{l(2)00681}$ eggs derived from GLCs. The presence of several background bands of equal intensity in the two different lanes reveals that similar amounts of proteins were loaded. Using four different affinity purified sera we have not been able to obtain specific *in situ* staining of the Ttv protein (see Methods).

Figure 4. Ttv is a type II integral membrane protein.

a. The Kyte-Doolittle hydrophobicity plot of the Ttv protein sequence shows the existence of a hydrophobic stretch at the N-terminus of the protein (aa 7 to 24, see arrow).

b. *In vitro* translation and microsome assays. In vitro transcribed Ttv cDNAs are translated into an 80kD protein in the presence of rabbit reticulocyte extract supplemented with 10 μCi of ^{35}S-methionine (lane 2). When the same extract is supplemented with dog pancreatic microsomes, Ttv is detected as a higher migrating band (lane 3). This shift can be abolished by endoglycosydase H treatment, indicating that the Ttv protein is glycosylated in the presence of microsomes (lane 4). The full length Ttv protein enters microsomes since Ttv is protected from proteinase K digestion in the presence of intact microsomes (lanes 5 and 6). Ttv remains associated with the membrane fraction after alkaline wash (lane 7 and 8), indicating that Ttv is a membrane protein.

c. Proposed structure for the Ttv protein. Note that the cytoplasmic tail is only six amino acids in length and is identical in Ext-1 and Ttv. We do not favor the attachment of Ttv to the membrane by a glycoyl-phophatidylinositol (GPI) anchor since GPI anchor proteins are characterized by a large hydrophobic region at the C-terminus (Low, M.G. Biochim Biophys Acta. The glycosyl-phosphatidylinositol anchor of membrane proteins. 988, 427-454 (1989)) not present in Ttv.

Figure 5. Scheme illustrating the generation of the typical domain structure of a HS chain during polysaccharide biosynthesis.

The symbols used are defined by the structures illustrated below the scheme. The open circle attached to the internal GlcNSO3 unit of the antithrombin-binding pentasaccharide sequence (AT) denotes a 3-O-sulfate group. The target residues for the various enzymes (NDST, GlcA C-5 epimerase (EPI), 2-OST, 6-OST, and 3-OST) are indicated by arrows and by red residue symbols. The arrows within parentheses indicate (arbitrary) sites of variable polymer modification, i.e. residues that satisfy the substrate specificity of the indicated reactions but nevertheless escape target selection. Two NDST isoforms are indicated, along with their potential target residues. For further information regarding isoforms of other enzymes (not indicated in the scheme) see the text. The reducing terminus is to the right.

Figures 6A-C: Ttv is involved in HSPGs biosynthesis

Staining of wild type (A-C) and *ttv* mutant (D-F) embryos with the 3G10 antibody. No staining was detected by 3G10 in embryos untreated with heparinase III (A and D). In embryos treated with heparinase III, uniform staining is detected at early stages (B) by 3G10. At later stages, stronger staining is observed in the nervous system (C). B and C are stage 10 and 13 embryos, respectively. In *ttv* embryos (E), the staining detected by 3G10 is strongly reduced. The decrease staining is not due to variability in staining conditions because when *ttv* is ectopically expressed in a *hairy* pattern in *ttv* embryos, the strong 3G10 staining is recovered in the *hairy* striped pattern.

Cuticles of wild type (G), *ttv* embryos (H), and *ttv/Df (2R)Trix* embryos derived from females with *ttv* germline clones (I). (J) Western blots of wild type (wt) and *ttv* embryos probed with rabbit anti-Ttv antibody. The predicted Molecular Weight of Ttv protein is 80 kD and a band comigrating with the 83 kD marker is detected in wild type and not in *ttv* embryos.

Figure 7: *Ext* protein comparison

Amino acid sequence comparison between Ttv, Homo sapiens Ext1 (Hs Ext1), *Drosophila* melanogaster Ext2 (Dm Ext2), and Homo sapiens Ext2 (Hs Ext2). Dm Ext2 is more similar to Hs Ext2 and Mm Ext2 than Ttv.

Figure 8: Subcellular localization of Ttv

Embryos stained for (A) Ttv-myc in green, (B) the Golgi marker ß-CopII in red, and (C) the ER marker Bip in blue. (D)Colocalization of Ttvmyc with ß-CopII is seen in yellow and with Bip in turquoise.

(E) Apical and (F) basolateral confocal sections of wild type embryo expressing the UAS-*ttvmyc* construct in the *hairy* domain. In these panels, Myc staining is shown in green and Hh in red. (G) Apical and (H) basolateral confocal section of wild type wing imaginal disc expressing *UAS-ttvmyc* in the *Ptc-Gal4* domain. In these panels, Myc staining is shown in red, and Ptc in green.

Figure 9: Ttv activity is not required for FGF signaling

Mesoderm migration visualized by Twi staining in wild type (A and C) and *ttv* embryos (B and D) at stage 9. A and C are ventral views of whole embryos, C and D are

transverse sections. In wild type cells, Twi stained cells can be seen as a band with smooth edge on the ventral side of stage 10 embryos and cross sections through these embryos show a monolayer of Twi-stained cells. When FGF signaling is impaired and mesodermal migration is affected, Twi positive cells can be seen as a band with a rough edge on the ventral side and sections through these embryos will show the cells clustered near the ventral midline.

Figure 10: Ttv activity is not required for Wg signaling

Wg staining in (A) wild type and (B) *ttv* embryo at stage 10. En staining in (C) wild type and (D) *ttv* embryo at stage 10. SNS was visualized by staining with anti-Crumbs antibodies in: (E) wild type, (F) *wg*, and (G) *ttv* embryos. Wing imaginal discs which are (H) wild type or (I) have a large clone of *ttv* mutant cells stained with Dll to assay for Wg diffusion. Wg staining in the epidermis of (J) wild type and (K) *ttv* mutant embryos.

Figure 11: Ttv affects Hh signaling and diffusion in the embryo

Expression of *bap* by RNA *in situ* of wild type (A and B) and *ttv* embryos (C and D). A and B are lateral views, C and D ventral views. Hh staining of endogenous Hh in wild type (E) and *ttv* embryos (F). (G) Staining of HhN in *ttv* embryos expressing *HhN* under control of *en* promoter (*en-GAL4 ; UAS-HhN*). (H) *ttv* embryos expressing *en-HhN* stained for both Hh in green and En in red.

Figures 12A and 12B: Models

(A) HS-GAGs are synthesized by a complex that includes Ttv and specific GAG-modifying enzyme(s) such as a sulfotransferases and C-5 epimerases. Different Ext complexes might include specific sulfotransferases not found in complex with other Ext proteins.

(B) Possible roles of HSPGs in Hh movement (see Example 5).

Detailed Description of the Invention

(i) Overview

Of particular importance in the development and maintenance of tissue in vertebrate animals is a type of extracellular communication called induction, which occurs between neighboring cell layers and tissues. In inductive interactions, chemical signals secreted by

one cell population influence the developmental fate of a second cell population. Typically, cells responding to the inductive signals are diverted from one cell fate to another, neither of which is the same as the fate of the signaling cells.

Glycosaminoglycans are sugars composed of long unbranched chains of repeating disaccharides, sometimes reaching hundreds of residues (reviewed in Kjellen, L. and Lindahl, U. (1991) Annu. Rev. Biochem. 60:443-475). These sugar chains are usually linked to a core protein forming a proteoglycan. There are two types of UDP-GlcDH-dependent glycosaminoglycans. The glucosaminoglycans, such as heparin and heparan sulfate, are synthesized as [GlcA β1,4- GlcNAc α1,4-$]_n$; the galactosaminoglycans, such as chondroitin sulfate, are synthesized as [GlcA β1,3- GalNAc α1,4-$]_n$. During chain elongation, the sugars are modified and extensively sulfated, giving the chains a strong negative charge. Different families within each chain type are generated by varying the kinds of modifications and the extent of sulfation.

Numerous studies have demonstrated that proteoglycans can regulate growth factor and cytokine activity *in vitro*. Typically, the ligand binds to the glycosaminoglycan chains on cell surface proteoglycans. This interaction can be quite strong; for example, the K_d for fibroblast growth factor (FGF) binding to syndecan, a heparan sulfated proteoglycan, is 10^{-9} M (Moscatelli, D. (1987) J. Cell. Physiol. 131:123-130; Klagsbrun, M. And Baird, A. (1991) Cell 67:229-231). Given the affinity of many growth factors for proteoglycans and the relative abundance of the latter, it is likely that most secreted growth factors are not freely diffusible but are, instead, sequestered on proteoglycans.

The present invention is directed to the discovery that there exists proteoglycans having specific glycosaminoglycan chains which provide exquisite selectivity for binding to particular growth factors or cytokines, e.g., growth factor-specific GAG chains. That is, there exists in cells proteoglycans that,by virtue of the particular modifications to the GAG chains of the protein, selectively bind one factor (or family or related factors). Moreover, the specificity of the GAG chain structure, which gives rise to this binding specificity, is regulated by a biosynthetic pathway which can be selectively inhibited in order to antagonize the activity of a specific growth factor.

To illustrate, one aspect of the present invention concerns the discovery that members of the Ext-family are required for Hh signaling. In particular, in order to determine the effect of the Ext-family gene, e.g., Ttv, on the diffusion of Hh, the inventors

analyzed the function of Ttv in the wing imaginal disc of Drosophila. In the wing imaginal disc, Hh is expressed in the cells of the posterior compartment and diffuse to the anterior compartment. Such diffusion may be detected by detecting Patched (Ptc) expression or Cubitus interruptus (Ci) stabilization, which are known targets of Hh signaling. It was observed that Hh is unable to diffuse in the absence of Ttv activity. As described in the appended examples, the Ext genes encode a family of glycosyltransferases. These enzymes synthesize GAG chains attached to the protein core of proteoglycans. The specificity of Ttv (a Drosophila Ext homolog) to hedgehog signaling suggests the existence of hedgehog-specific GAG chains. To account for the specificity of Ttv to hedgehog signaling, as well as the existence of multiple Ext genes without apparent redundancy or tissue specificity, we propose that different Ext glycosyltransferances form specific complexes in the Golgi with different GAG modifying enzymes, such as sulfotransferases, deacetylases and the like. Each complex can generate qualitatively different GAG sequences with unique specificity for binding cytokines and other serum proteins.

Example 6 (below) further illustrates the invention. Characterization of the *Drosophila* gene *sulfateless*, which encodes a homolog of a vertebrate heparan sulfate N-deacetylase/N-sulfotransferase (a member of a class of enzymes essential for the modification of heparan sulfate), reveals that HS proteoglycans are necessary for Wg/Wnt signaling. Moreover, we have identified Dally, a GPI-linked Glypican, as the HSPG molecule involved in Wg signaling. Loss of *dally* activity, both in the embryo and imaginal dics, generates phenotypes reminiscent to loss of Wg activity. Interestingly, *dally* is co-expressed with the Wg receptor frizzled 2 (Dfz2). In Drosophila, and higher eukaryotes, our data suggests that glypicans serve as co-receptors for the Wnt receptor, e.g., frizzled proteins, and together with frizzled proteins, modulate both short and long-range activities of Wnt proteins.

Based on the above understanding, one aspect of the present invention provides drug discovery assays for identifying agents which can selectively alter the activity of a growth factor, cytokine or other serum proteins (collectively referred to hereinafter as "factors" for ease of reading). In particular, the subject assay is designed to identify inhibitors of specific modifications of GAG chains, such as particular enzymes involved in, e.g. isomerization, sulfation, or acetylation which produce factor-specific binding by the resulting proteoglycan. Such inhibitors can be used to alter the specificity of a proteoglycan for binding a particular factor (or family of related factors) by inhibiting

specific modifications to the sugar sequence of the proteoglycan. In one embodiment, the assay identifies agents which selectively inhibit a sulfotransferase, preferably an N-deacetylase/N-sulfotransferase, involved in synthesis of a proteoglycan that binds a selected factor - so as to decrease the ability of the proteoglycan to bind the factor.

Another aspect of the present invention relates to a method for inhibiting the activity of a factor, *in vitro* or *in vivo*, by administering a compound which selectively inhibits synthesis of the factor-specific glycosaminoglycan in an amount sufficient to reduce the responsiveness of the treated cell(s) to the factor. Such methods can be used to treat, e.g., unwanted cell proliferation or other unwanted effects of a factor.

Yet another aspect of the present invention provides a method for potentiating the activity of a selected factor, *in vitro* or *in vivo*, by administering a compound which selectively inhibits synthesis of GAG chains which bind other factors that compete with the selected factor, without substantially diminishing the synthesis of GAG chains which bind the selected factor. In such a manner, the activity of the factor can be potentiated, e.g., its mitogenic, trophic or other activity enhanced.

In one embodiment, this invention provides drug discovery assays for identifying agents which can modulate the biological function of selected factors, such as by (i) inhibiting the glycosyltransferase activity of an Ext protein, e.g. Ext1; (ii) inhibiting a GAG-modifying activity of an enzyme associated with an Ext protein, e.g. a sulfotransferase; or (iii) inhibiting the interaction of an Ext protein with other enzymes involved in GAG synthesis, e.g. sulfotransferases. For example, the subject method can be used to inhibit hedgehog-mediated signal transduction by inhibiting synthesis of hedgehog specific GAG chains, e.g., by Ext-dependent pathways. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a tissue otherwise responsive to hedgehog proteins. For example, the subject inhibitors can be used to prevent hedgehog-dependent proliferation of cells, such as hedgehog-dependent basal cell carcinoma (BCC) or other hedgehog-dependent epithelia proliferative disorders. In another example, the subject inhibitors can be used to alter the growth state of a mesodermally-derived tissue, such as cartilage. In particular, this invention provides a method for identifying agents that are effective in amielorating at least one symptom of a disorder characterized by aberrant bone morphogenesis, such as multiple exostoses, chondrosarcomas, and/or osteosarcomas. In other embodiments, the invention also helps in identifying agents useful therapeutically to inhibit hedgehog-mediated growth,

maintenance and/or differentiation of a tissue involved in spermatogenesis and tissue derived from dorsal mesoderm; ectodermally-derived tissue, such as tissue derived from the epidermis, neural tube, neural crest, or head mesenchyme; endodermally-derived tissue, such as tissue derived from the primitive gut.

In another embodiment, this invention provides drug discovery assays for identifying agents which can modulate the biological function of selected factors which bind to frizzled receptors, such as Wnt proteins. For example, the assays can be used to identify inhibitors of Wnt activity by (i) inhibiting Wnt-specfic modification of glypicans by a GAG-modifying activity, e.g. inhibiting a glypican-modifying N-deacetylase/N-sulfotransferase; (ii) inhibiting the interaction of a Wnt protein with a frizzled/glypican complex; or (iii) inhibiting formation of a a frizzled-glypican complex capable of binding to a Wnt protein. Accordingly, the subject method can be used to inhibit Wnt-mediated signal transduction by inhibiting synthesis of Wnt-specific GAG chains, e.g., present on Wnt-specific glypicans. Such agents can be useful therapeutically to alter the growth, maintenance and/or differentiation of a tissue otherwise responsive to Wnt proteins. The identification of modulators of Wnt activity are useful in treating disease states involving the Wnt activity. Such compounds could be of use in the treatment of diseases in which activation or inactivation of the Wnt protein results in either cellular proliferation, cell death, nonproliferation, induction of cellular neoplastic transformations or metastatic tumor growth and hence could be used in the prevention and/or treatment of cancers such as bone and breast cancer for example.

In yet another embodiment, the present invention provides diagnostic and therapeutic assays and reagents for detecting and treating disorders involving, for example, aberrant expression (or loss thereof) of an Ext glycosyltransferase gene, abberant expression of a *dally* N-deacetylase/N-sulfotransferase gene, or abberant expression of a glypican.

(ii) Definitions

For convenience, certain terms employed in the specification and appended claims are collected here.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made

from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a polypeptide, including both exon and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a polypeptide, e.g., comprising coding exon sequences and (optionally) intron sequences. The term "intron" refers to a DNA sequence present in a given gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a polypeptide or, where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the protein is disrupted.

As used herein, the term "specifically hybridizes" refers to the ability of a nucleic acid probe/primer of the invention to hybridize to at least 15 consecutive nucleotides of a gene, such as a *Ext* sequence designated in any one or more of SEQ ID Nos: 1-3 or a sequence complementary thereto, or naturally occurring mutants thereof, such that it has less than 15%, preferably less than 10%, and more preferably less than 5% background hybridization to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a protein, as defined herein.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However,

the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant *Ext* gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring forms of *Ext* genes.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of neuronal or hematopoietic origin. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but can cause at least low level expression in other tissues as well.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding a first polypeptide with a second amino acid sequence defining a domain (e.g. polypeptide portion) foreign to and not substantially homologous with any domain of the firsst protein. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula X-*P*-Y, wherein *P* represents a portion of the fusion protein which is derived from a first protein, and X and Y are, independently, absent or represent amino acid sequences which are not related to the P sequences in an organism.

The term "substantially homologous", when used in connection with amino acid sequences, refers to sequences which are substantially identical to or similar in sequence, giving rise to a homology in conformation and thus to similar biological activity. The term is not intended to imply a common evolution of the sequences.

The term "percent identical" refers to sequence identity between two amino acid sequences or between two nucleotide sequences. Identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When an equivalent position in the compared sequences is occupied by the same base or amino acid, then the molecules are identical at that position; when the equivalent site occupied by the same or a similar amino acid residue (e.g., similar in steric and/or electronic nature), then the molecules can be referred to as homologous (similar) at that position. Expression as a percentage of homology/similarity or identity refers to a function of the number of identical or similar amino acids at positions shared by the compared sequences. Various alignment algorithms and/or programs may be used, including FASTA, BLAST or ENTREZ. FASTA and BLAST are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with, e.g., default settings. ENTREZ is available through the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences.

The term "ortholog" refers to genes or proteins which are homologs via speciation, e.g., closely related and assumed to have common descent based on structural and functional considerations. Orthologous proteins function as recognizably the same activity in different species.

The term "paralog" refers to genes or proteins which are homologs via gene duplication, e.g., duplicated variants of a gene within a genome. See also, Fritch, WM (1970) *Syst Zool* 19:99-113.

The term "Ext-family protein" or "Ext-family polypeptide" refers to a family of polypeptides characterized at least in part by being identical or sharing a degree of sequence homology with all or a portion of the Ext-family polypeptides represented in any of SEQ ID Nos: 4-6. The Ext-family polypeptides can be cloned or purified from any of a number of invertebrate or vertebrate organisms, especially vertebrates, and particularly mammals. Moreover, other Ext-family polypeptides can be generated according to the present invention, which polypeptides do not ordinarily exist in nature, but rather are generated by non-natural mutagenic techniques.

As used herein, the term "vertebrate *hedgehog* protein" refers to vertebrate inter-cellular signaling molecules related to the *drosophila* hedgehog protein. Three of the vertebrate *hedgehog* proteins, *Desert hedgehog* (*Dhh*), *Sonic hedgehog* (*Shh*) and *Indian hedgehog* (*Ihh*), apparently exist in all vertebrates, including amphibians, fish, birds, and mammals. Other members of this family, such as *Banded hedgehog, Cephalic hedgehog, tiggy-winkle hedgehog*, and *echidna hedgehog* have been so far identified in fish and/or amphibians. Exemplary *hedgehog* polypeptides are described in PCT applications WO96/17924, WO96/16668, WO95/18856.

The terms " Wnts" or "Wnt gene product" or " Wnt polypeptide" when used herein encompass native sequence Wnt polypeptides, Wnt polypeptide variants, Wnt polypeptide fragments and chimeric Wnt polypeptides. The definition specifically includes human Wnt polypeptides, Wnt -1, 2, 3a, 3b, , 4, 5a, 5b, 6, 7a, 7b, 8a, 8b, 10a, 10b, 11, and 12.

As used herein, the terms "transforming growth factor-beta" and "TGF- " denote a family of structurally related paracrine polypeptides found ubiquitously in vertebrates, and prototypic of a large family of metazoan growth, differentiation, and morphogenesis factors (see, for review, Massaque et al. (1990) *Ann Rev Cell Biol* 6:597-641; and Sporn et al. (1992) *J Cell Biol* 119:1017-1021). Included in this family are the "bone morphogenetic proteins" or "BMPs", which refers to proteins isolated from bone, and fragments thereof and synthetic peptides which are capable of inducing bone deposition alone or when combined with appropriate cofactors. Preparation of BMPs, such as BMP-1, -2, -3, and -4, is described in, for example, PCT publication WO 88/00205. Wozney (1989) *Growth Fact Res* 1:267-280 describes additional BMP proteins closely related to BMP-2, and which have been designated BMP-5, -6, and -7. PCT publications WO89/09787 and WO89/09788 describe a protein called "OP-1," now known to be BMP-7. Other BMPs are known in the art.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

As used herein, "phenotype" refers to the entire physical, biochemical, and physiological makeup of a cell, e.g., having any one trait or any group of traits.

By "enhancing differentiation of a cell" is meant the act of increasing the extent of the acquisition or possession of one or more characteristics or functions which differ from that of the original cell (i.e., cell specialization). This can be detected by screening for a change in the phenotype of the cell (e.g., identifying morphological changes in the cell and/or surface markers on the cell).

By "enhancing survival or maintenance of a cell" encompasses the step of increasing the extent of the possession of one or more characteristics or functions which are the same as that of the original cell (i.e., cell phenotype maintenance).

The terms "induction" or "induce", as relating to the biological activity of a growth factor, cytokine or other factor, refers generally to the process or act of causing to occur a specific effect on the phenotype of cell. Such effect can be in the form of causing a change in the phenotype, e.g., proliferation, differentiation to another cell phenotype, secretion, or can be in the form of maintaining the cell, e.g., preventing dedifferentation or promoting survival of a cell.

As used herein, the term "target tissue" refers to connective tissue, cartilage, bone tissue or limb tissue, which is either present in an animal, e.g., a mammal, e.g., a human or is present in in vitro culture, e.g., a cell culture.

An "effective amount" of an agent, with respect to the subject method of treatment, refers to an amount of agonist or antagonist in a preparation which, when applied as part of a desired dosage regimen, provides modulation of growth, differentiation or survival of cells, e.g., modulation of skeletogenesis, e.g., osteogenesis, chondrogenesis, or limb patterning, spermatogenesis, and neuronal differentiation.

A "patient" or "subject" to be treated can mean either a human or non-human animal.

The term “small organic molecule” refers to a non-peptide, non-nucleotide organic compound having a molecular weight less than 7500amu, more preferably less than 2500amu, and even more preferably less than 750amu.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter gene" operatively linked to a transcriptional regulatory sequences. Transcription of the reporter gene is controlled by these sequences. The activity of at least one or more of these control sequences is directly or indirectly regulated by a signal transduction pathway involving a phospholipase, e.g., is directly or indirectly regulated by a second messenger produced by the phospholipase activity. The transcriptional regulatory sequences can

include a promoter and other regulatory regions, such as enhancer sequences, that modulate the activity of the promoter, or regulatory sequences that modulate the activity or efficiency of the RNA polymerase that recognizes the promoter, or regulatory sequences that are recognized by effector molecules, including those that are specifically induced upon activation of a phospholipase. For example, modulation of the activity of the promoter may be effected by altering the RNA polymerase binding to the promoter region, or, alternatively, by interfering with initiation of transcription or elongation of the mRNA. Such sequences are herein collectively referred to as transcriptional regulatory elements or sequences. In addition, the construct may include sequences of nucleotides that alter the stability or rate of translation of the resulting mRNA in response to second messages, thereby altering the amount of reporter gene product.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding a polypeptide preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks the gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes*

(IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

(iii) Illustrative Embodiments

A. Drug-Screening Assays

The present invention provides drug screening assays for identifying pharmaceutically effective compounds that specifically potentiate or inhibit the activity of a growth factor, cytokine or other serum protein. One aspect of the present invention is directed to methods for screening compounds which inhibit the synthesis of factor-selective glycosaminoglycan chains, e.g., by inhibiting one or more glycosyltransferases which are essential and/or specific for synthesis of a factor selective GAG. A "factor-selective GAG" is one which confers a selective binding activity to a proteoglycan for a selected factor. Thus, for example, a factor-selective GAG can be characterisexed by it ability to confer a K_d for binding to a selected factor (or family of related factors) which is at least one order of magnitude, more preferably two, three or four orders of magnitude less than the K_d for other factors which are unrelated to the selected factor. In other embodiments, the factor-selectivity of the GAG chain may be characterized by its ability to confer an IC_{50} or EC_{50} for an activity of the selected factor which is at least one order of magnitude, more preferably two, three or four orders of magnitude less than the IC_{50} or EC_{50} for other factors which are unrelated to the selected factor or relative to the absence of the specific GAG chain. An "unrelated factor" is one which does not share sufficient homology or biological activity to be considered an ortholog or paralog of the selected factor, e.g., it may be less than 50, 60, 70, 80 or 90 percent identical to the selected factor.

In certain illustrative embodiments, the assay can be formated to identify an inhibitor of a glycosyltransferase(s) such as, for example: an N-acetylgalactosaminyltransferase (GalNAc transferase) such as EC 2.4.1.40, EC 2.4.1.92,

EC 2.4.1.41; an N-acetylglucosaminyltransferase (GlcNAc transferase) such as EC 2.4.1.101, EC 2.4.1.143, EC 2.4.1.144, EC 2.4.1.155; an N-acetylglucosaminyl-1-phosphate transferase (GlcNAc-1-P transferase) such as EC 2.7.8.15; a Fucosyltransferase (Fuc transferase) such as EC 2.4.1.69, EC 2.4.1.65; a galactosyltransferase (Gal transferase) such as EC 2.4.1.151, EC 2.4.1.90, EC 2.4.1.22, EC 2.4.1.37, EC 2.4.1.45; a Glucosyltransferase (Glc transferase) such as EC 2.4.1.117; a glucuronyltransferase such as EC 2.4.1.17; a Mannosyltransferase (Man transferase) such as EC 2.4.1.83, EC 2.4.1.109; an oligosaccharyltransferase such as EC 2.4.1.119; a sialyltransferase (NeuAc transferase) such as EC 2.4.99.4, EC 2.4.99.6, EC 2.4.99.1, EC 2.4.99.3, EC 2.4.99.8. Ext1 and Ext2 are similar to glucuronyl/N-acetylglucosaminyl transferase

In one preferred embodiment, the assay is formated to identify inhibitors of one or more glycosaminoglycan sulfotransferases or glycosaminoglycan deacetylases. Glycosaminoglycan N-deacetylases/N-sulfotransferases are structurally related enzymes that play an important role in the biosynthesis of heparan sulfate and heparin. They are dual catalytic, single membrane-spanning polypeptides of approximately 850-880 amino acids that catalyze the N-deacetylation of N-acetylglucosamine of glycosaminoglycans followed by N-sulfation of the same sugar. These enzymes play a pivotal role in the biosynthesis of heparan sulfate and heparin. Such an enzyme was first purified and cloned from rat liver (Brandan et al. (1988) J. Biol. Chem. 263:2417; Hashimoto et al. (1992) J. Biol. Chem. 267:15744), and subsequently isoforms were cloned from mouse mastocytoma cells (Orellana et al. (1994) J. Biol. Chem. 269:2270), the mouse tumor (Eriksson et al. (1994) J. Biol. Chem. 269:10438), as well as human tissues (Gladwin et al. (1996) Genomics 32:471). Experiments in vitro and in situ have shown that a single polypeptide chain of approximately 900 amino acids mediates the above two reactions without any additional proteins or cofactors (Wei et al. (1993) PNAS 90:3885; Mandon et al. (1994) J. Biol. Chem. 269:11729; and Orellana et al., supra).

In preferred embodiments, the sulfotransferase is selected from amongst the EC (Enzyme Commission) sulfotransferase sub-subclass 2.8.2.-, e.g., an aryl sulfotransferase (EC 2.8.2.1), an arylamine sulfotransferase (EC 2.8.2.3), a chondroitin 4-sulfotransferase (EC 2.8.2.5), a UDP-N-acetylgalactosamine-4-sulfate sulfotransferase (EC 2.8.2.7), a desulfoheparin sulfotransferase (EC 2.8.2.8), a galactosylceramide sulfotransferase (EC 2.8.2.11), a heparitin sulfotransferase (EC 2.8.2.12), a chondroitin 6-sulfotransferase (EC 2.8.2.17), a triglucosylalkylacylglycerol sulfotransferase (EC 2.8.2.19), a keratan

sulfotransferase (EC 2.8.2.21), an arylsulfate sulfotransferase (EC 2.8.2.22), a heparin-glucosamine 3-O-sulfotransferase (EC 2.8.2.23), or a desulfoglucosinolate sulfotransferase (EC 2.8.2.24).

In a preferred embodiment, the glycosyltransferase is an N-deacetylase (EC 3.1.1.-) / N-sulfotransferase (EC 2.8.2.-).

In a preferred embodiment, the glycosyltransferase is sulfotransferase which interacts with an Ext protein and/or is essential for hedgehog-specific GAG chains.

In a preferred embodiment, the glycosyltransferase is sulfotransferase which is essential for Wnt-specific GAG chains, e.g., is essential for synthesis of glypicans which selectively interact with Wnt proteins.

Exemplary sulfotransferases which can be used in the subject assay are described in Tables 1 and 2. The subject assay can also be carried out using related homologs, e.g., which are at least 75, 85, 90 or 95 percent identical to one of the sulfotransferases enumerated herein, or which hybridize under under stringent to highly stringent hybridization and wash conditions (e.g., 2 x SSC at 50°C to 0.2 x SSC at 50°C).

Table 1

Enzyme	**Reference**
Chondroitin-6-sulphotransferase	Fukuta et al (1995) J. Biol. Chem. 270:18575-18580
Glycosaminoglycan N-acetylglucosamine N-deacetylase/N-sulphotransferase 1	Dixon et al (1995) Genomics 26:239-241 Hashimoto et al (1992) J. Biol. Chem. 267:15744-15750
Glycosaminoglycan N-acetylglucosamine N-deacetylase/N-sulphotransferase 2	Orellana et al (1994) J. Biol. Chem. 269:2270-2276 Eriksson et al (1994) J. Biol. Chem. 269:10438-10443

Table 2

Enzyme	**NCBI ENTREZ Accession***
heparan sulfate n-deacetylase/n-sulfotransferase 3 [Homo sapiens]	4322247

n-deacetylase/n-sulfotransferase (heparan glucosaminyl) [Homo sapiens]	4505353
n-deacetylase/n-sulfotransferase (heparan glucosaminyl) [Homo sapiens]	4505351
heparan sulfate n-deacetylase/n-sulfotransferase 1 [Mus musculus]	4322251
Heparan sulfate n-deacetylase/n-sulfotransferase (n-hsst)	1708322
heparan sulfate n-deacetylase/n-sulfotransferase 2 [Mus musculus]	4322249
heparin/heparan sulfate n-acetylglucosaminyl n-deacetylase/n-sulfotransferase [Bos taurus]	3885496
heparan sulfate glucosaminyl n-deacetylase/n-sulfotransferase; dast-1 [Mus musculus]	3136148
heparan glucosaminyl n-deacetylase/n-sulfotransferase-2 [Homo sapiens]	2792518
heparan sulfate-N-deacetylase/N-sulfotransferase - human	2135335
Heparin sulfate n-deacetylase/n-sulfotransferase (n-hsst)	1708324
Heparin sulfate n-deacetylase/n-sulfotransferase (n-hsst)	1708323
heparan n-deacetylase/n-sulfotransferase-2 [Homo sapiens]	1036799
heparan n-deacetylase/n-sulfotransferase-1 [Homo sapiens]	1036797
heparan sulfate-N-deacetylase/N-sulfotransferase [Homo sapiens]	976372
heparan sulfate N-deacetylase/N-sulfotransferase [Homo sapiens]	841164

mastocytoma glucosaminyl N-deacetylase/N sulfotransferase - mouse	627909
glycosaminoglycan N-acetylglucosaminyl N-deacetylase (EC 3.1.1.-) / N-sulfotransferase (EC 2.8.2.-) - mouse	539823
glucosaminyl n-deacetylase [Mus musculus]	474431
glycosaminoglycan N-acetylglucosaminyl N-deacetylase/N-sulfotransferase [Mus musculus]	457944
N-heparan sulfate sulfotransferase - rat	348494
N-heparan sulfate sulfotransferase [Rattus norvegicus]	205703
heparan sulfate-N-deacetylase/N-sulfotransferase - human	2135335

*The Entrez Browser is provided by the National Center for Biotechnology Information

The subject assay can be carried out by any of a variety of drug screening techniques. Such assays can be used to identify agents, especially small organic molecules (preferably <750 amu) which either potentiate or inhibit the glycosyltransferase activity of one or more enzymes involved in synthesis of a factor-selective GAG chain. The enzymes, or fragments thereof, may either be free in solution, affixed to a solid support, or provided in a host cell. For each of the assay embodiments set out herein, the assay is preferably repeated for a variegated library of at least 100 different test compounds, though preferably libraries of at least 10^3, 10^5, 10^7, and 10^9 compunds are tested. The test compound can be, for example, small organic molecules, and/or natural product extracts.

In an exemplary cell-free embodiment, a selected glycosyltransferase is isolated in solution, e.g., as a reconsitituted (or otherwise purified) protein preparation or from a cell-lysate. The enzyme is contacted with an appopriate substrate and a test compound, and rate of the enzyme-mediated conversion of the substrate to product is assessed and compared to the rate of conversion in the absense of the test compound.

In the case of an N-deacetylase (EC 3.1.1.-) / N-sulfotransferase, a variety of assays have been developed in the art to detect such activities, and can be readily adapted for use in the subject methods.

In one embodiment, N-deacetylase activity can be measured by determining the release of [^{3}H]acetate from N-[^{3}H]acetylated-polysaccharide, e.g., isolated from E. coli K5-derived capsular polysaccharide as described (Bame et al., (1991) J. Biol. Chem. 266:12461-12468,). N-sulfotransferase and O-sulfotransferase activities can be measured by determining $^{35}SO_4$-incorporation into N-desulfated heparin and completely desulfated, N-resulfated heparin, respectively, from ^{35}S-labeled adenosine 3'phosphate phosphosulfate (PAPS) as described (Ishihara et al., (1992) Anal. Biochem. 206:400-407).

Another assay format for sulfotransferase activity which can be readily adapted for use in the subject drug screening assays is descibed by http://www.panvera.com/tech/protocols/p2435lit.html. Briefly, the assay uses the following reagents

Dilution Buffer: 5 mM potassium phosphate buffer, pH 6.5, containing 1.5 mg/ml BSA and 10 mM dithiothreitol.

Cocktail Buffer: (for one reaction only): 25 µl of 50 mM potassium phosphate buffer, pH 6.5, 25 µl of dithiothreitol (7.4 mg/ml) and 1.28 mM 35S-PAPS (NEN, NEG-010).

Stop Mixture: a 1:1 vol/vol mixture of 0.1 M $Ba(OH)_2$ and 0.1 M barium acetate made fresh each day.

The steps are as follows:

1. Place 10 µl of substrate/additive into a 1.5 ml microfuge tube and leave on ice.
2. Add 100 µl of enzyme in dilution buffer.
3. Start reaction by addition of 50 µl of "cocktail" and test agent (though test agent may be added in step 2).
4. Incubate at 37°C for 20 minutes.
5. Stop the reaction by addition of 100 µl of the "stop mixture", vortex. Then add 50 µl 0.1 M $ZnSO_4$ and vortex again. You should see a precipitate form.
6. Centrifuge for 3 minutes at maximum speed in a microfuge.
7. Add 50 µl 0.1 M $Ba(OH)_2$ to the tube, vortex. Then add 50 µl 0.1 M $ZnSO_4$ and vortex again.
8. Centrifuge for 10 minutes at maximum speed in a microfuge and aliquot 300 µl of the supernatant into a scintillation vial. Add 5 ml of scintillant and count.

Sulfotransferases can show some degree of overlapping substrate specificity. It is therefore important to perform experiments under linear conditions and define the kinetic

parameters for substrate conjugation before making conclusions about isozyme specificity. Also these enzymes can show significant substrate inhibition, and it may be necessary to try a variety of substrate concentrations.

In another illustrative embodiment, a cell-free assay for UDP-glucuronic acid (Singh et al. (1980) Biochem. J. 189:369) can be utilized for determining UDP-glucose dehydrogenase activity. The assay for UDP-glucuronic acid, a product of UDP-glucose dehydrogenase, is based on the fluorometric determination of D-glucuronosyl benzo(a)pyrene. This compound is formed from UDP-glucuronic acid and 3-hydroxybenzo(a)pyrene in a reaction catalyzed by the glycuronosyl transferase of guinea pig microsomes. Unreacted 3-hydroxybenzo(a)pyrene is removed by extraction with chloroform-methanol, and the amount of gluconosylbenzo(a)pyrene formed is determined fluorometrically. See also Burrows et al. (1983) Anal Biochem 130:376-8.

Davies et al. (1972) Anal Biochem 47):209-17 describes a radiochemical assay for UDP-glucose dehydrogenase which can be used in the subject drug screening assays.

In still other embodiments, the drug screening assay can be carried out in a cell-based format. In a preferred embodiment, a cell (e.g., a eukaryotic cell) includes a reporter gene which is regulated by the selected factor, e.g., its expression is induced or repressed by hedgehog, a Wnt or other growth factor or cytokine. The cell is preferably engineered to express one or more of, a proteoglycan which can specifically interact with the selected factor, and/or a glycosyltransferase which is essential for generating a factor-selective GAG chain. For example, the cell can be engineered to express a recombinant EXT or sulfotransferase which interacts therewith, or a glypican or sulfotransferase which modifies the GAG chain(s) of the glypican. The cell is contacted with a test agent(s), and the level of expression of the reporter gene is assessed. An change in the level of expression of the reporter gene, relative to the control of no test agent, indicates that the test agent is able to modulate (potentiate or inhibit) the activity of selected factor.

In still other embodiments, the invention provides drug screening assays which are capable of identifying agents that can disrupt the glypican-dependent interaction between a Wnt protein it cognate receptor, e.g., with a frizzled protein, or which inhibit the formation of glypican-frizzled receptor complexes. An exemplary drug screening assay of the present invention includes the steps of (a) forming a reaction mixture including: (i) a Wnt polypeptide, (ii) glypican-frizzled complex, and (iii) a test compound; and (b) detecting interaction of the Wnt polypeptide with the complex. A statistically significant change

(potentiation or inhibition) in the interaction of the Wnt polypeptide with the complex in the presence of the test compound, relative to the interaction in the absence of the test compound, indicates a potential agonist (mimetic or potentiator) or antagonist (inhibitor) of Wnt bioactivity for the test compound. The reaction mixture can be a cell-free protein preparation, e.g., a reconstituted protein mixture or a cell lysate, or it can be a recombinant cell including a heterologous nucleic acid recombinantly expressing the glypican polypeptide.

Where the drug screening assay is a cell-free system, it can be, e.g., a cell membrane preparation, a reconstituted protein mixture, or a liposome reconstituting the receptor subunits as a Wnt receptor. For instance, the protein subunits of a Wnt receptor complex (including the glypican component) can be purified from detergent extracts from both authentic and recombinant origins can be reconstituted in artificial lipid vesicles (e.g. phosphatidylcholine liposomes) or in cell membrane-derived vesicles (see, for example, Bear et al. (1992) Cell 68:809-818; Newton et al. (1983) Biochemistry 22:6110-6117; and Reber et al. (1987) J Biol Chem 262:11369-11374). The lamellar structure and size of the resulting liposomes can be characterized using electron microscopy. External orientation of the receptor in the reconstituted membranes can be demonstrated, for example, by immunoelectron microscopy. The interaction of a Wnt protein with liposomes containing such glypican-frizzled complexes and liposomes without the protein, in the presence of candidate agents, can be compared in order to identify potential modulators of the interaction.

In yet another embodiment, the drug screening assay is derived to include a whole cell expressing a recombinant glypican polypeptide. The ability of a test agent to alter the activity of the glypican-frizzled complex can be detected by analysis of the recombinant cell. For example, agonists and antagonists of the receptors biological activity can by detected by scoring for alterations in growth or differentiation (phenotype) of the cell. General techniques for detecting each are well known, and will vary with respect to the source of the particular reagent cell utilized in any given assay. For the cell-based assays, the recombinant cell is preferably a metazoan cell, e.g., a mammalian cell, e.g., an insect cell, e.g., a xenopus cell, e.g., an oocyte. In other embodiments, the receptor can be reconstituted in a yeast cell.

In an exemplary embodiment, a cell which expresses the receeombinant receptor, can be contacted with a Wnt protein which is capable of inducing signal transduction from

the receptor, and the resulting signaling detected either at various points in the pathway, or on the basis of a phenotypic change to the reagent cell. A test compound which modulates that pathway, e.g., potentiates or inhibits, can be detected by comparison with control experiments which either lack the receptor or lack the test compound. For example, visual inspection of the morphology of the reagent cell can be used to determine whether the biological activity of the glypican-dependent Wnt pathway has been affected by the added agent.

In addition to morphological studies, change(s) in the level of an intracellular second messenger responsive to signaling by the glypican complex can be detected. For example, in various embodiments the assay may assess the ability of test agent to cause changes in phosphorylation patterns, adenylate cyclase activity (cAMP production), GTP hydrolysis, calcium mobilization, and/or phospholipid hydrolysis (IP_3, DAG production) upon receptor stimulation. A reporter gene can also be provided in the cell, being selected to include a transcriptional regulatory element which is responsive to glypican-dependent Wnt signal transduction.

Simple competitive binding assays can also be used to assess the ability of a test compound to inhibit formation of glypican-Wnt complexes in cell-free mixtures.

After identifying certain test compounds as being active in one of the subject assays, the practitioner of the subject assay will continue to test the efficacy and specificity of the selected compounds both *in vitro* and *in vivo*. Whether for subsequent *in vivo* testing, or for administration to an animal as an approved drug, agents identified in the subject assay can be formulated in pharmaceutical preparations for *in vivo* administration to an animal, preferably a human.

B. Sources of Genes and Proteins

It is contemplated by the present invention that the cloned Ext-family genes set out in the appended sequence listing, in addition to representing an inter-species family of related genes, are also each part of an intra-species family. That is, it is anticipated that other paralogs of the human and Drosophila Ext-family proteins exist in those animals, and orthologs of each Ext-family gene are conserved amongst other animals.

According to the appended sequence listing, (see also Table 3) human Ext-1 is a polypeptide encoded by a nucleic acid represented by SEQ ID No:1; human Ext-2 is encoded by the nucleic acid SEQ ID No: 2; Ttv polypeptide is encoded by SEQ ID No: 3.

Table 3
Guide to Ext-family sequences in Sequence Listing

	Nucleotide	Amino Acid
Human Ext-1	SEQ ID No. 1	SEQ ID No. 4
Human *Ext-2*	SEQ ID No. 2	SEQ ID No. 5
Drosophila Ttv	SEQ ID No. 3	SEQ ID No. 6

A human Ext-1 gene has been localized to chromosomal position 8q24.1, human Ext-2 (Gen. Bank Accession No. U62740) has been localized to 11p11-13, and human Ext-3 has been localized to 19p . Other Ext family members include, but are not limited to ExtR1 (Gen. Bank Accession No. AB007042), ExtR2 (Gen. Bank Accession No. AB009284), Ext2.I (Gen. Bank Accession No. U72263), ExtL (Gen. Bank Accession No. U67191), ExtL3 (Gen. Bank Accession No. U76188) and ExtL2 (Gen. Bank Accession No. U76189).

The coding sequence for the Drosophile *dally* protein can found as GenBank accession U31985, and is also described by Nakato et al. (1995) Development 121:3687. The drosophila *dally* protein is closely related to the mammalian glypican proteins, particularly human glypican 5. The coding sequence for human glypican 5 is provided herein as SEQ IS No. 7, and the corresponding protein sequence as SEQ ID No. 8. Other exemplary mammalian glypicans, or related proteins, which can be used in the subject assays are set forth in Table 4

Table 4

Enzyme	**NCBI ENTREZ Accession***
glypican 1 [Homo sapiens]	4504081
glypican-1 precursor	1707999
glypican-1 precursor	462190
heparan sulfate proteoglycan	227480
glypican 1 precursor - human	106224
glypican [Homo sapiens]	31847
glypican 4 [Homo sapiens]	4504083
glypican-4 [Homo sapiens]	3831547
glypican 4 [Homo sapiens]	3420277
heparan sulfate proteoglycan - rat	2143780

k-glypican [Mus musculus]	758627
glypican [Rattus norvegicus]	506417
heparan sulphate proteoglycan [Homo sapiens]	1245417
glypican [Homo sapiens]	1237181
GTR2-2 gene product [Homo sapiens]	976205
mxr7 [Homo sapiens]	556647
intestinal protein OCI-5 [Rattus norvegicus]	205800
development-specific protein OCI-5 - rat	111579
cerebroglycan	740477

*The Entrez Browser is provided by the National Center for Biotechnology Information

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a Ext, a glypican, or a glycosyltransferase essential to synthesis of a factor-selective GAG chain, so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a the targeted protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775), or peptide nucleic acids (PNAs). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by

Van der Krol et al. (1988) Biotechniques 6:958-976; and Stein et al. (1988) Cancer Res 48:2659-2668.

C. Diagnostic Applications

The present method also provides a method for determining if a subject is at risk for a disorder characterized by aberrant apoptosis, cell proliferation and/or differentiation. In preferred embodiments, method can be generally characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding a *Ext*-protein, or (ii) the mis-expression of the *Ext* gene or *glypican* gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a *Ext* gene or *glypican* gene, (ii) an addition of one or more nucleotides to a *Ext* gene or *glypican* gene, (iii) a substitution of one or more nucleotides of a *Ext* gene or *glypican* gene, (iv) a gross chromosomal rearrangement of a *Ext* gene or *glypican* gene, (v) a gross alteration in the level of a messenger RNA transcript of a *Ext* gene or *glypican* gene, (vii) aberrant modification of a *Ext* gene or *glypican* gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a *Ext* gene or *glypican* gene, (viii) a non-wild type level of a *Ext*-protein, and (ix) inappropriate post-translational modification of a *Ext*-protein. As set out below, the present invention provides a large number of assay techniques for detecting lesions in a *Ext* gene or *glypican* gene, and importantly, provides the ability to discern between different molecular causes underlying *Ext*-dependent aberrant cell growth, proliferation and/or differentiation.

In an exemplary embodiment, there is provided a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a *Ext* gene or *glypican* gene, such as represented by any one of SEQ ID Nos: 1-4 or 7, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject *Ext* gene or *glypican* genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA

level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels.

In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1944) PNAS 91:360-364), the later of which can be particularly useful for detecting point mutations in the *Ext* gene or *glypican* gene. In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to a *Ext* gene or *glypican* gene under conditions such that hybridization and amplification of the *Ext* gene or *glypican* gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample.

In still another embodiment, such assays as described above can be used to detect mutations, e.g., loss-of-function mutations, to a gene encoding a glycoyltransferase which is essential to signal transduction by a growth factor, cytokine or the like. To illustrate, the human homolog of *sulfateless*, see example, 6, is a heparan sulfate-N-deacetylase/N-sulfotransferase. The human gene maps to 5q32-q33.3. That locus is associated with such developmental disorders as Treacher Collins syndrome. Treacher Collins syndrome is an autosomal dominant disorder of craniofacial development, which has been localized to chromosome 5q32-33.1. The symptoms of the disorder are reminiscent of abberant signal transduction by a Wnt or hedgehog protein.

D. Method of Treating Diseases

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or inhibiting (or alternatively potentiating) proliferation of a cell, by contacting the cells with an agent which modulates Hedgehog- or Wnt-dependent signal transduction pathways. The subject method could be used to generate and/or maintain an array of different tissue both *in vitro* and *in vivo*. A "Hedgehog or Wnt therapeutic," whether inhibitory or potentiating with respect to modulating the activity of a Hedgehog or Wnt protein respectively. can be, as appropriate, any of the preparations described above, including agents identified in the subject assays,

and antisense molecules which selectively inhibits expression of a glycosltransferase enzyme, such as an Ext or a sulfotransferase, essential to the synthesis of hedgehog-selective or Wnt-selective GAG chains, or an antisense molecule which selectively inhibits expression of a glypican.

The Hedgehog therapeutic compounds of the present invention are likely to play an important role in the modulation of cellular proliferation and maintenance of, for example, neuronal, testicular, osteogenic or chondrogenic tissues during disease states. Accordingly, in one embodiment of the present invention, compositions comprising Hedgehog or Wnt therapeutics can be used in the *in vitro* generation of skeletal tissue, such as from skeletogenic stem cells, as well as the *in vivo* treatment of skeletal tissue deficiencies. Hedgehog therapeutics are also expected to play an important role in regulating uncontrolled cellular proliferation leading to exostoses, and in severe cases progression of the exostoses to Chondrosarcomas and Osteosarcomas.

For instance, it is known in the art that proper regulation of chondrocyte differentiation is necessary for the morphogenesis of skeletal elements, yet little is known about the molecular regulation of this process. Indian hedgehog (Ihh) is expressed in the prehypertrophic chondrocytes of cartilage elements, where it regulates the rate of hypertrophic differentiation. Misexpression of Ihh prevents proliferating chondrocytes from initiating the hypertrophic differentiation process. The direct target of Ihh signaling is the perichondrium where Gli and Ptc flank the expression domain of Ihh. Ihh induces the expression of a second signal, parathyroid hormone-related protein (PTHrP), in the periarticular perichondrium. Analysis of PTHrP (-/-) mutant mice indicated that the PTHrP protein signals to its receptor in the prehypertrophic chondrocytes, thereby blocking hypertrophic differentiation. In vitro application of Hedgehog or PTHrP mediates the effects of Ihh through the formation of a negative feedback loop that modulates the rate of chondrocyte differentiation.

As proliferating chondrocytes decide to undergo hypertrophy, they express high levels of the PTH/PTHrP receptor. When they subsequently become committed to this pathway, they transiently express Ihh, until they become fully hypertrophic. The Ihh signal acts on the perichondrium adjacent to the prehypertrophic zone (Ptc/Gli-expressing cells) and (directly or indirectly) on the more distant periarticular perichondrium. ultimately inducing the expression of PTHrP. PTHrP then signals back to chondrocytes expressing

the PTH/PTHrP receptor, thereby preventing non-differentiated chondrocytes from moving down the hypertrophic pathway.

Accordingly, the present invention particularly contemplates the use of Hedgehog therapeutics which agonize and/or antagonize a *hedgehog*, e.g., Ihh activity as the case may be.

In a particular embodiment, this invention contemplates the use of Hedgehog therapeutics, in alleviating the symptoms associated with multiple exostoses. Hereditary multiple exostoses is an autosomal dominant condition characterized by multiple cartilage-capped projections originating from the epiphyseal plates of endochondral bones (Jaffe 1943; Solomon 1961). The prevalence of this disorder has been estimated at between 1/50,000 and 1/100,000 (Hennekam 1991; Cook et al. 1993; Schmale et al. 1994). The development of exostoses appears to be correlated with growth of the epiphyseal plate, with new exostoses rarely appearing after epiphyseal plate closure (Hennekam 1991). Exostoses usually form at the ends of long bones but can also occur on the hands, feet, pelvis, ribs, and scapula. Shortened stature is sometimes seen in this disorder and is possibly attributable to these benign tumors interfering with the epiphyseal growth plates and reducing overall bone growth (Hennekam 1991). Exostoses also frequently impinge on adjacent nerves, vessels, tendons, and joints and thus require surgical removal. A serious complication of this disorder is the occasional transformation to chondrosarcoma or osteosarcoma that occurs in 0.9%-2.9% of cases (Voutsinas and Wynne-Davies 1983; Wicklund et al. 1995).

Genetic linkage analysis in families with multiple exostoses has already identified three loci responsible for hereditary multiple exostoses on chromosomes 8q24.1 (Ext1), 11p11-13 (Ext2), and 19p (Ext3) (Cook et al. 1993; Le Merrer et al. 1994; Wu et al. 1994; Wuyts et al. 1995). To date, the exostoses that are caused by disruption of these genes appear to be indistinguishable; no specific phenotypic differences have been observed in patients from different linkage groups (Le Merrer et al. 1994). As discussed above, the role of Ext-family proteins in hedgehog signaling indicates that multiple exostoses is associated with abnormal diffusion of Hh proteins. Hedgehog therapeutics that agonize Ihh activity and modulate diffusion of Ihh would therefore be useful in alleviating the symptoms associated with Multiple exostoses.

In another embodiment, the present invention contemplates the use of Hedgehog therapeutics which agonize a hedgehog skeletogenic activity, such as an ability to induce

chondrogenesis and/or osteogenesis. By "skeletal tissue deficiency", it is meant a deficiency in bone or other skeletal connective tissue at any site where it is desired to restore the bone or connective tissue, no matter how the deficiency originated, e.g. whether as a result of surgical intervention, removal of tumor, ulceration, implant, fracture, or other traumatic or degenerative conditions. In addition, the present invention makes available effective therapeutic methods and compositions for restoring cartilage function to a connective tissue. Such methods are useful in, for example, the repair of defects or lesions in cartilage tissue which is the result of degenerative wear such as that which results in arthritis, as well as other mechanical derangements which may be caused by trauma to the tissue, such as a displacement of torn meniscus tissue, meniscectomy, a laxation of a joint by a torn ligament, malignment of joints, bone fracture, or by hereditary disease. The present reparative method is also useful for remodeling cartilage matrix, such as in plastic or reconstructive surgery, as well as periodontal surgery. The present method may also be applied to improving a previous reparative procedure, for example, following surgical repair of a meniscus, ligament, or cartilage. Furthermore, it may prevent the onset or exacerbation of degenerative disease if applied early enough after trauma.

In one embodiment of the present invention, the subject method comprises treating the afflicted connective tissue with a therapeutically sufficient amount of a *hedgehog* agonist, particularly Hedgehog therapeutic which agonizes *Ihh* activity, to generate a cartilage repair response in the connective tissue by stimulating the differentiation and/or proliferation of chondrocytes embedded in the tissue. Induction of chondrocytes by treatment with a *hedgehog* agonist can subsequently result in the synthesis of new cartilage matrix by the treated cells. Such connective tissues as articular cartilage, interarticular cartilage (menisci), costal cartilage (connecting the true ribs and the sternum), ligaments, and tendons are particularly amenable to treatment in reconstructive and/or regenerative therapies using the subject method. As used herein, regenerative therapies include treatment of degenerative states which have progressed to the point of which impairment of the tissue is obviously manifest, as well as preventive treatments of tissue where degeneration is in its earliest stages or imminent. The subject method can further be used to prevent the spread of mineralisation into fibrotic tissue by maintaining a constant production of new cartilage.

In an illustrative embodiment, the subject method can be used to treat cartilage of a diarthroidal joint, such as a knee, an ankle, an elbow, a wrist, a knuckle of either a finger or

toe, or a temperomandibular joint. The treatment can be directed to the meniscus of the joint, to the articular cartilage of the joint, or both. To further illustrate, the subject method can be used to treat a degenerative disorder of a knee, such as which might be the result of traumatic injury (e.g., a sports injury or excessive wear) or osteoarthritis. An injection of a Hedgehog therapeutic into the joint with, for instance, an arthroscopic needle, can be used to treat the afflicted cartilage. In some instances, the injected agent can be in the form of a hydrogel or other slow release vehicle described above in order to permit a more hedgehog or extended and regular contact of the agent with the treated tissue.

The present invention further contemplates the use of the subject method in the field of cartilage transplantation and prosthetic device therapies. To date, the growth of new cartilage from either transplantation of autologous or allogenic cartilage has been largely unsuccessful. Problems arise, for instance, because the characteristics of cartilage and fibrocartilage varies between different tissue: such as between articular, meniscal cartilage, ligaments, and tendons, between the two ends of the same ligament or tendon, and between the superficial and deep parts of the tissue. The zonal arrangement of these tissues may reflect a gradual change in mechanical properties, and failure occurs when implanted tissue, which has not differentiated under those conditions, lacks the ability to appropriately respond. For instance, when meniscal cartilage is used to repair anterior cruciate ligaments, the tissue undergoes a metaplasia to pure fibrous tissue. By promoting chondrogenesis, the subject method can be used to particularly addresses this problem, by causing the implanted cells to become more adaptive to the new environment and effectively resemble hypertrophic chondrocytes of an earlier developmental stage of the tissue. Thus, the action of chondrogenesis in the implanted tissue, as provided by the subject method, and the mechanical forces on the actively remodeling tissue can synergize to produce an improved implant more suitable for the new function to which it is to be put.

In similar fashion, the subject method can be applied to enhancing both the generation of prosthetic cartilage devices and to their implantation. The need for improved treatment has motivated research aimed at creating new cartilage that is based on collagen-glycosaminoglycan templates (Stone et al. (1990) *Clin Orthop Relat Red* 252:129), isolated chondrocytes (Grande et al. (1989) *J Orthop Res* 7:208; and Takigawa et al. (1987) *Bone Miner* 2:449), and chondrocytes attached to natural or synthetic polymers (Walitani et al. (1989) *J Bone Jt Surg* 71B:74; Vacanti et al. (1991) *Plast Reconstr Surg* 88:753; von Schroeder et al. (1991) *J Biomed Mater Res* 25:329; Freed et al. (1993) *J Biomed Mater*

Res 27:11; and the Vacanti et al. U.S. Patent No. 5,041,138). For example, chondrocytes can be grown in culture on biodegradable, biocompatible highly porous scaffolds formed from polymers such as polyglycolic acid, polylactic acid, agarose gel, or other polymers which degrade over time as function of hydrolysis of the polymer backbone into innocuous monomers. The matrices are designed to allow adequate nutrient and gas exchange to the cells until engraftment occurs. The cells can be cultured *in vitro* until adequate cell volume and density has developed for the cells to be implanted. One advantage of the matrices is that they can be cast or molded into a desired shape on an individual basis, so that the final product closely resembles the patient's own ear or nose (by way of example), or flexible matrices can be used which allow for manipulation at the time of implantation, as in a joint.

In one embodiment of the subject method, the implants are contacted with a Hedgehog therapeutic during the culturing process, such as an *Ihh* agonist, in order to induce and/or maintain differentiated chondrocytes in the culture in order as to further stimulate cartilage matrix production within the implant. In such a manner, the cultured cells can be caused to maintain a phenotype typical of a chondrogenic cell (i.e. hypertrophic), and hence continue the population of the matrix and production of cartilage tissue.

In another embodiment, the implanted device is treated with a Hedgehog therapeutic in order to actively remodel the implanted matrix and to make it more suitable for its intended function. As set out above with respect to tissue transplants, the artificial transplants suffer from the same deficiency of not being derived in a setting which is comparable to the actual mechanical environment in which the matrix is implanted. The activation of the chondrocytes in the matrix by the subject method can allow the implant to acquire characteristics similar to the tissue for which it is intended to replace.

In yet another embodiment, the subject method is used to enhance attachment of prosthetic devices. To illustrate, the subject method can be used in the implantation of a periodontal prosthesis, wherein the treatment of the surrounding connective tissue stimulates formation of periodontal ligament about the prosthesis, as well as inhibits formation of fibrotic tissue proximate the prosthetic device.

In still further embodiments, the subject method can be employed for the generation of bone (osteogenesis) at a site in the animal where such skeletal tissue is deficient. Indian *hedgehog* is particularly associated with the hypertrophic chondrocytes that are ultimately

replaced by osteoblasts. For instance, administration of a Hedgehog or Wnt therapeutic of the present invention can be employed as part of a method for treating bone loss in a subject, e.g. to prevent and/or reverse osteoporosis and other osteopenic disorders, as well as to regulate bone growth and maturation. For example, preparations comprising *hedgehog* agonists can be employed, for example, to induce endochondral ossification, at least so far as to facilitate the formation of cartilaginous tissue precursors to form the "model" for ossification. Therapeutic compositions of Hedgehog or Wnt therapeutics can be supplemented, if required, with other osteoinductive factors, such as bone growth factors (e.g. TGF- factors, such as the bone morphogenetic factors *BMP-2* and *BMP-4*, as well as activin), and may also include, or be administered in combination with, an inhibitor of bone resorption such as estrogen, bisphosphonate, sodium fluoride, calcitonin, or tamoxifen, or related compounds. However, it will be appreciated that *hedgehog* proteins, such as I*hh* and S*hh* are likely to be upstream of BMPs, e.g. treatment with a *hedgehog* agonist will have the advantage of initiating endogenous expression of BMPs along with other factors.

It will also be apparent that, by transient use of modulators of Hedgehog activities, *in vivo* reformation of tissue can be accomplished, e.g. in the development and maintenance of organs such as ectodermal patterning, as well as certain mesodermal and endodermal differentiation processes. By controlling the proliferative and differentiative potential for different cells, the subject Hedgehog therapeutics can be used to reform injured tissue, or to improve grafting and morphology of transplanted tissue. For instance, Hedgehog antagonists and agonists can be employed in a differential manner to regulate different stages of organ repair after physical, chemical or pathological insult. The present method is also applicable to cell culture techniques.

In another embodiment, the subject hedgehog inhibitors of the present invention can be used to inhibit hedgehog-mediated proliferation of cells, e.g., to treat unwanted proliferation. In one embodiment, the subject inhibitors can be used to treat hedgehog-dependent basal cell carcinoma (BCC) or other cancer resulting from over-activation of hedgehog signaling.

As described above, in another embodiment the subject invention provides inhibitors of Wnt signal transduction. Such compounds could be of use in the treatment of diseases in which activation or inactivation of the Wnt protein results in either cellular proliferation, cell death, nonproliferation, induction of cellular neoplastic transformations

or metastatic tumor growth and hence could be used in the prevention and/or treatment of cancers such as bone and breast cancer for example.

All of the above-cited references and publications are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention.

Example 1: Ext-family polypeptides affect Hh Signaling

We identified mutations in a novel gene, *tout velu* (*ttv*), in a large screen for maternal effect mutations associated with segment polarity phenotypes. (Perrimon, N., Lanjuin, A., Arnold, C. & Noll, E. Gene*tics* Zygotic lethal mutations with maternal effect phenotypes in Drosophila melanogaster. II. Loci on the second and third chromosomes identified by Pelement-induced mutations. 144, 1681-1692 (1996)). The segment polarity phenotype is reminiscent of the phenotype of *wingless* (*wg*) or *hedgehog* (*hh*) embryos. Since several specific *wg*-dependent processes are not affected in embryos that lack both maternal and zygotic *ttv* gene activities, it is likely that *ttv* is required for Hh signaling (I.T., Y.B. and N.P., in preparation). Consistent with this hypothesis, *ttv* somatic clones are associated with wing phenotypes similar to the phenotypes of clones of mutations in Hh signaling components (data not shown). In this paper we analyze the function of *ttv* in detail in the wing imaginal disc to determine its role in the Hh signaling pathway.

In the wing disc, Hh is expressed in cells of the posterior (P) compartment and has been proposed to diffuse into the anterior (A) compartment. Within the anterior compartment, binding of Hh to its receptor Patched (Ptc) activates the Hh signaling pathway that results in the stabilization of the full length Cubitus interruptus (Ci) protein, a transcription factor of the GLI family. (Tabata, T., Eaton, S. & Kornberg, T.B. Gene*s Dev* - The Drosophila hedgehog gene is expressed specifically in posterior compartment cells and is a target of engrailed regulation. 6, 2635-2645 (1992)); (Basler, K. & Struhl, G. *Nature*

Compartment boundaries and the control of Drosophila hmb pattern by hedgehog protein. 368, 208-214 (1994)); (Marigo, V., Davey, R.A., Zuo, Y., Cunningham, J.M. & Tabin, C.J. *Nature* Biochemical evidence that patched is the Hedgehog receptor. 384, 176179 (1996)); (Stone, D.M., *et al.* *Nature* The tumor-suppressor gene patched encodes a candidate receptor for Sonic hedgehog., 129-134 (1996)); (Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C. & Kornberg, T.B. *Cell* Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. 89, 1043-1053 (1997)); (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)); and (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)). Ci then activates the expression of *ptc*, which restricts the range of diffusion of Hh. (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)). Consistent with the model that Hh diffuses into the A compartment, Ptc expression can be detected as far as 5 cell diameters away from the antero-posterior (A/P) boundary and the stabilization of Ci can be detected as far as 8 to 10 cell diameters (Figures la, lb, and see also (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997))). We therefore used Ptc expression and Ci stabilization as reporters to examine if *ttv* is involved in Hh signaling.

In large *ttv* mutant clones adjacent to the A/P border, within the anterior compartment, Hh signaling is impaired as concluded from the lack of Ptc staining in most of the clone (Figures lc, le). However, Ptc expression is still induced at the posterior edge of the clone adjacent to the wild type cells (Figures le). The level of signaling in these cells is nevertheless diminished, as shown by a reduction in their levels of Ptc staining compared to wild type cells (Figure le). A similar result was observed when Ci levels were used as a reporter for Hh signaling (Figures lf, lh). In addition, the large domain of Ci stabilization allowed us to determine that a cell non autonomous effect was associated with *ttv* mutant clones. When *ttv* mutant clones of only a few cells wide are located along the A/P boundary, wild type cells anterior to the *ttv* mutant clone do not respond to the Hh signal as shown by Ci staining (Figures li, lk). Because these cells are located within a domain where Hh is normally able to diffuse to and induce stabilization of Ci, we conclude that *ttv* mutant clones have a directional non cell autonomous effect on wild type cells located anteriorly. Since the domain of Ci stabilization is under the direct control of Hh

signaling (Aza-Blanc, P., Ramirez-Weber, F.A., Laget, M.P., Schwartz, C. & Kornberg, T.B. *Cell* Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. 89, 1043-1053 (1997)); (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)), we interpret this result as the inability of Hh to reach wild type cells located anterior to *ttv* mutant clones. This observation led us to investigate if the diffusion of Hh is impaired in absence of *ttv*.

Example 2: Ext-family Polypeptides are Required for Hh Diffusion

To analyze whether *ttv* is required for Hh diffusion, we took advantage of the fact that Hh can diffuse over a greater distance in the absence of Ptc. (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)). In a *ptc* mutant clone, Hh diffusion is observed as an ectopic induction of Ptc in wild type cells localized anterior to a clone of *ptc* mutant cells (Figure 2a and (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)). To determine if Hh diffusion is modified by *ttv* in the *ptc* mutant clones, we generated double mutant clones for *ptc* and *ttv* and analyzed the expression of Ptc in wild type cells anterior to the clones. Interestingly, in *ptc ttv* double mutant clones Ptc expression is not induced in wild type cells anterior to the clone (Figure 2b).

To directly assess Hh diffusion in the *ptc ttv* mutant clones, we compared the distribution of Hh in *ptc* versus *ptc ttv* clones. In the *ptc* mutant cells, we detect the Hh protein as a diffuse membrane staining. When Hh reaches the wild type cells beyond a *ptc* clone, it can be seen in a punctate staining pattern that for the most part coincides with the punctate Ptc staining (Figure 2c). In *ptc ttv* double mutant clones, we do not detect any Hh staining in mutant cells (Figure 2d). Based on this result and the directional cell non-autonomous effect of *ttv* mutant clones, we propose that Hh is unable to diffuse in the absence of *ttv* activity.

In order to diffuse Hh has to move from the sending to the receiving cell, therefore *ttv* could function in the sending cell and/or the receiving cell. Importantly, there are two types of sending cells, the posterior cells that produce and send Hh and the anterior cells which send previously received Hh. Since in the anterior cell the sending of Hh depends on its reception from the previous cells, it is difficult to address where *ttv* functions by generating *ttv* mutant clones in the anterior compartment. We therefore generated clones in

the posterior compartment adjacent to the A/P boundary and analyzed their effects on Ci stabilization. We observed that *ttv* mutant clones in the posterior compartment do not affect the diffusion of Hh since the stabilization of Ci is not affected in the anterior compartment (Figures 2e, 2and see also 2b). Therefore, we propose that *ttv* functions in the receiving cells and is probably required for the movement of Hh from sending to receiving cells. This observation also indicates that *ttv* is not required for Hh production.

Interestingly, we observed that in a *ttv* clone located in the anterior compartment, Hh can still signal although less efficiently to the first receiving cells. We hypothesize that this weakened signaling activity is mediated by the Hh present on the membrane of the Hh sending cells and that therefore efficient Hh signaling even to adjacent cells requires Hh diffusion.

Example 3: Cloning of Ext-family Member

To understand the molecular function of *ttv* in Hh diffusion, we cloned the *ttv* gene. We isolated the genomic region surrounding the *ttv$^{l(2)00681}$* P-element insertion (Figure 3a) and we identified two genes in the vicinity of the P-element insertion site. One is *lamin* C, which is unlikely to correspond to *ttv* since it is not expressed maternally. (Riemer, D., *et al. J Cell Sci* Expression of Drosophila lamin C Is developmentally regulated: analogies with vertebrate A-type lamins. 108, 3189-3198 (1995)). The other gene is maternally expressed and we isolated its full length cDNA. Sequencing of the putative *ttv* 3.8 kb cDNA reveals that it encodes a 760 amino acid protein which is similar to the human Multiple Exostoses (Ext-1) protein (Ahn, J., *et al. Nat* Gene*t* Cloning of the putative tumor suppressor gene for hereditary multiple exostoses (EXT1). 11, 137-143 (1995)); (Figures 3b and 3c). The 3.8 kb transcript identifies *ttv* since a transgene of this transcript under the control of a *hsp*7O promoter rescues *ttv* homozygous flies to viability. The *ttv$^{l(2)00681}$* mutation is either a null or a severe hypomorphic allele consistent with the genetic data (see methods), since the wild type protein is not detected in protein extracts made from unfertilized eggs derived from *ttv$^{l(2)00681}$* germline clone females (Figure 3d).

Ttv is 56% and 25% identical to the human Ext-1 and *Ext-2* protein, respectively. The *Ext-1* and *Ext-2* genes have implicated in the human Multiple Exostoses syndrome.. (Ahn, J., *et al. Nat* Gene*t* Cloning of the putative tumor suppressor gene for hereditary

multiple exostoses (EXT1). 11, 137-143 (1995)); (Stickens, D., *et al. Nat Genet* The EXT2 multiple exostoses gene defines a family of putative tumor suppressor genes. 14, 25-32 (1996)) and (Wicklund, C.L., Pauli, R.M., Johnston, D. & Hecht, J.T. *Am j Med Genet*. Natural history study of hereditary multiple exostoses. 55, 43-46 (1995)). Moreover, the Ext proteins appear to form a novel family of conserved molecules in Metazoans, since mouse and C. elegans homologues have been cloned recently. (Clines, G.A., Ashley, J.A., Shah, S. & Lovett, M. *Genome Res* The structure of the human multiple exostoses 2 gene and characterization of homologs in mouse and Caenorhabditis elegans. 7, 359-367 (1997)). However, so far, no studies have determined the cellular compartment where Ext molecules could function. (Ahn, J., *et al. Nat Genet* Cloning of the putative tumor suppressor gene for hereditary multiple exostoses (EXT1). 11, 137-143 (1995)) and (Stickens, D., *et al. Nat Genet* The EXT2 multiple exostoses gene defines a family of putative tumor suppressor genes. 14, 25-32 (1996)). To determine more precisely how Ttv mediates the movement of Hh, we investigated this issue. Analysis of the Ttv sequence reveals the presence of a hydrophobic stretch at the N-terminus of the Ttv protein which is conserved in the human Ext-1 and Ext-2 proteins, indicating that *ttv* might code for a transmembrane protein (Figure 4a). We used an *in vitro* translation assay to demonstrate that Ttv is indeed a membrane protein (Figure 4b). *In vitro* transcribed *ttv* mRNA is translated into a protein of approximately 80 kD, which is in accordance with its calculated molecular weight. In the presence of microsomes, the protein is glycosylated and fully protected from degradation by proteinase K, clearly indicating that the protein is imported into microsomes. Furthermore, the Ttv protein remains associated with the membrane fraction following alkaline wash, demonstrating that it is a membrane associated protein. (Fujiki, Y., Hubbard, A.L., Fowler, S. & Lazarow, P.B. *J Cell Biol* .Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. 93, 97-102 (1982)). Since the sequence surrounding the signal sequence does not have a good consensus for cleavage (von Heijne, G. *Nucleic Acids Res* .A new method for predicting signal sequence cleavage sites. 14, 4683-90 (1986)), we propose that the signal sequence is not cleaved but rather acts as an anchor region. From this analysis, we conclude that Ttv is a type II integral membrane protein (Figure 4c).

Example 4: Characteristics of the Ext-Polypeptide

Our observation that Ttv is an integral membrane protein might provide a first clue to reconcile a puzzle in understanding Hh diffusion. In contrast to most signaling ligands, the active form of Hh is tethered to the membrane by a cholesterol moiety, (Porter, J.A., Young, K.E. & Beachy, P.A. *Science* Cholesterol modification of hedgehog signaling proteins in animal development. 274, 255-259 (1996)) and (Porter, J.A., *et al. Cell* Hedgehog patteming activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. 86, 21-34 (1996)), thus it is not clear how Hh can diffuse from one cell to another. The tethering of Hh suggests that Hh diffusion is mediated by a regulated process at the cell surface that allows Hh to move from a sending to a receiving cell. We therefore imagine that Hh should first be dissociated from the sending cell membrane and then re-associated with the receiving cell membrane. The movement of Hh cannot be dependent on the Hh receptor Ptc and the membrane signaling component Smoothened (Smo), since Hh diffuses in *ptc* or *smo* mutant clones (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)). We therefore propose the existence of a Smo, Ptc independent process that regulates the diffusion of Hh signaling molecule. Interestingly, Ttv appears to play a important role in this process. Since Ttv is required in the receiving cell to allow Hh diffusion, we speculate that Ttv does not function in the dissociation of Hh from the membrane of the sending cells but more likely to permit the reassociation or the maintenance of Hh on the surface of the receiving cells.

Until now, due to the absence of homology between Ext and other known proteins, the function of the Ext family of molecules has remained unknown. As a consequence, no model has been proposed to explain the pathophysiology associated with the Multiple Exostose syndrome. The Multiple Exostoses syndrome is a dominantly inherited disease characterized by short stature, limb length inequalities, bone deformities and the presence of bone outgrowths, called exostoses, at the ends of long bones (Wicklund, C.L., Pauli, R.M., Johnston, D. & Hecht, J.T. *Am j Med Genet.* Natural history study of hereditary multiple exostoses. 55, 43-46 (1995)). The loss of any one of the *Ext-1, 2* or *3* locus has been shown to be responsible for this syndrome (Wicklund, C.L., Pauli, R.M., Johnston, D. & Hecht, J.T. *Am j Med Genet.* Natural history study of hereditary multiple exostoses. 55, 43-46 (1995)). The *Ext-1* and *Ext-2* genes have recently been cloned (Ahn, J., *et al. Nat Genet* Cloning of the putative tumor suppressor gene for hereditary multiple exostoses (EXT1). 11, 137-143 (1995)) and (Stickens, D., *et al. Nat Genet* The EXT2 multiple

exostoses gene defines a family of putative tumor suppressor genes. 14, 25-32 (1996)) and are homologous to each other. *Ext-1* and *Ext-2* genes have been proposed to act as putative tumor suppressor genes due to the slightly elevated incidence of chondrosarcomas and osteosarcomas in Ext patients (Ahn, J., *et al. Nat* Genet Cloning of the putative tumor suppressor gene for hereditary multiple exostoses (EXT1). 11, 137-143 (1995)); (Stickens, D., *et al. Nat* Genet The EXT2 multiple exostoses gene defines a family of putative tumor suppressor genes. 14, 25-32 (1996)) and (Wicklund, C.L., Pauli, R.M., Johnston, D. & Hecht, J.T. *Am j Med* Genet. Natural history study of hereditary multiple exostoses. 55, 43-46 (1995)).

Interestingly, during bone morphogenesis Indian Hh is expressed in the prehypertrophic chondrocytes, and appears to signal to the adjacent tissue, the perichondrium where Ptc is expressed. Misexpression studies have clearly shown that Indian Hh (Ihh) limits the rate of chondrocyte differentiation by inducing the expression of the parathyroid hormone-related protein (PTHRP) in the perichondrium (Vortkamp, A., *et al. Science* Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. 273, 613-622 (1996)) and (Lanske B., *et al. Science* PTH/PTHRP receptor in early development and Indian hedgehog-regulated bone growth. 273, 663-666 (1996)). Based on the high similarity between Ext-1 and Ttv, we propose that some aspects of the syndrome, the short stature, limb length inequalities and bone deformities are due to defects in Ihh diffusion and efficient signaling. This hypothesis would be consistent with the model that Ihh regulates cartilage growth and therefore bone growth by limiting the rate of chondrocyte differentiation (Vortkamp, A., *et al. Science* Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. 273, 613-622 (1996)) and (Lanske B., *et al. Science* PTH/PTHRP receptor in early development and Indian hedgehog-regulated bone growth. 273, 663-666 (1996)). More recently, a careful analysis of the Ihh expression pattern strongly suggests that Ihh has a number of additional roles during bone morphogenesis (Vortkamp, A., *et al. Mechanism of development* Recapitulation of signals regulating embryonic bone formation during postnatal growth and in fracture repair. in press (1998)). We expect that the understanding of additional roles of Ihh during bone development will provide a mechanism to account for the appearance of exostoses or the slightly elevated incidence of chondrosarcoma in Ext patients.

A number of recent studies in vertebrates and invertebrates indicate that the patterning activity of Hh depends on its local concentration, (Johnson, R.L. & Tabin, C. *Cell.* The long and short of hedgehog signaling. 81, 313-316 (1995)); (Strigini, M. & Cohen, S.M. *Development* A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. 124, 4697-4705 (1997)); (Ericson, J., *et al. Cell* .Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. 90, 169-180 (1997)) and (Mullor, J., Calleja, M., Capdevila, J. & Guerrero, I. *Dev Stippl* .Hedgehog activity, independent of Decapentaplegic, participates in wing disc patterning. 124, 1227-1237 (1997)). The Hh receptor, Ptc, has recently been proposed to sequester Hh and therefore to increase Hh concentration close to the Hh source (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)). We propose that the control of the local concentration of Hh also depends on an opposite activity that require Ttv. We found that Ttv is required for the diffusion of Hh molecules and therefore increases the level of Hh far from its source. Moreover, *ttv* affects the diffusion of Hh to the first receiving cells and at further distance, suggesting that both the mechanisms of Hh short and long range diffusion depends on *ttv*, and are most likely identical. The characterization of *ttv* is an important step toward the understanding of a cellular mechanism that allows Hh to diffuse and to exert its patterning activity.

Methods

Somatic clonal analyses. The $ttv^{l(2)00681}$ allele behaves as a genetic null since the phenotype of embryos derived from ttvl(2)00681 germline clones (GLCS) (Low, M.G. *Biochim Biophys Acta* . The glycosyl-phosphatidylinositol anchor of membrane proteins. 988, 427-454 (1989)), crossed with $ttv^{l(2)00681}$*/Cyo* is identical to the phenotype of embryos derived form $ttv^{l(2)00681}$ GLC crossed with *Df(2R)Trix/Cyo*. S omatic clones were induced using the FRT/FLP mediated recombination system (Xu, T. & Rubin, G.M. Development Analysis of genetic mosaics in developing and adult Drosophila tissues. 117, 1223-1237 (1993)). Various strains were constructed for these analyses : for ttv clones: y w hsFLP; $FRT^{G13}ttv^{l(2)00681}/FRT^{G13}$ arm-LacZ,, for ptc clones: y w hsFLP;FR $T^{42D}ptc^{IW}$ / FRT^{42D} arm-lacZ, ptc^{IW} is a protein null allele (Chen, Y. & Struhl, G. *Cell* Dual roles for patched in sequestering and transducing Hedgehog. 87, 553-563 (1996)); and for ptc ttv double mutant clones: y w

hsFLP;FR $T^{42D}ptc^{IW}$ $ttv^{l(2)00681}$. First instar larvae were heat shocked for one hour at 37 C. Third instar larval imaginal discs were dissected in PBS and fixed for 20 minutes in 4% formaldehyde in PBS with 0.1% Tween2O (PBT). Antibodies were diluted in PBT as follows: rabbit anti- -Gal (Cappel) 1:4000; rat anti- -Gal 1:100; mouse anti-Ptc 1:100; rabbit anti-Hh 1:2000 (TSA amplification was used to enhance the signal, NEN); rat anti-Ci 1:5. Secondary antibodies coupled to FITC, Texas Red or Cy5 (Jackson) were used at 1:200 dilution. Secondary coupled to HRP for TSA amplification (Vector) was used at 1:2000 dilution. Stainings were visualized using a Leica TCS/NP confocal microscope.

Rescue construct: A SpeI, NotI fragment of the ttv gene was cloned downstream of the hsp7O promoter of the pCasper-hsp vector and used to transform the y w strain using the helper plasmid pp25-1. One insertion on the third chromosome rescued ttv homozygous mutants to viability without heat shock. $ttv^{l(2)00681}$ and two other alleles $ttv^{l(2)kO66109}$ and $ttv^{l(2)02055}$ were tested.

Antibody production : the ttv cDNA sequence encoding amino acid 30 to 376 of the Ttv protein was cloned into the pQE 10 vector (Qiagen) and expressed in the E.coli M15 strain. The purified protein was injected into rabbit and rat. Antisera were affinity purified against GST-Ttv (amino acid 30 to 376) linked on a HiTrap NHS-activated column (Pharmacia). No specific stainings could be detected with these affinity purified sera in embryos or imaginal discs.

Western Blot : Wild type and ttv unfertilized embryos derived from GLCs (Perrimon, N., Lanjuin, A., Arnold, C. & Noll, E. Gene*tics* Zygotic lethal mutations with maternal effect phenotypes in Drosophila melanogaster. II. Loci on the second and third chromosomes identified by Pelement-induced mutations. 144, 1681-1692 (1996)); were homogenized in RIPA buffer (50 mM Tris pH 8, 0. 1% SDS, 1 % NP40, 150 mM NACI)- 50µg of the extract was then separated by SDS PAGE electrophoresis (4% to 15% gradient gel) and blotted onto PVDF membrane (Millipore). Affinity purified rabbit anti-Ttv antibodies and secondary antirabbit-IG (vector) were used at 1:1000 and 1:3000 respectively. Detection by the ECL Western blotting detection method (Amersham).

In vitro translation assays: ttv RNA were produced by in vitro transcription using the SP6 RNA Machine kit (Ambion); rabbit reticulocyte lysates, microsomes and endoglycosydase H were purchased from Boehringer. In vitro translation assays in presence of 10 µCi ^{35}S methionine, Proteinase K digestion (0.lmg/ml per reaction) and endoglycosydase H treatments (2 mU/reaction) were performed according to manufacturer

protocols, except that lmM of PMSF was added at the end of the reaction. Proteins were then precipitated by adding 10 volumes of 66% saturated $(NH_4)_2SO_4$, resuspended in H_20 and reprecipitated with 10% TCA. The samples were then resuspended in sample buffer and analyzed by SDS PAGE. The alkaline wash was performed by adding 12 reaction volumes of. 0.1 M Na_2CO_3 (pH 12) to the sample and incubating it for 5 minutes at 4"C. The sample was then spinned at 75 000 rpm for 15 minutes. The supernatant was removed and the pellet was dissolved in loading buffer and analyzed by SDS PAGE.

Sequence analysis. The hydrophobic plot was generated using the Protean Software. The analysis of the peptide cleavage site was done on the PSORT II web server (http://psort.nibb.ac.jp:80/form2.html) using a algorithm derived from (von Heijne, G. *Nucleic Acids Res* .A new method for predicting signal sequence cleavage sites. 14, 4683-90 (1986)).

Standard molecular biology protocols were used for Northern blot, library screening and molecular cloning.

Example 5: Specific Heparan Sulfate Proteoglycans regulate the movement of Hedgehog in *Drosophila* tissues

We present evidence that the *Drosophila* gene *tout velu (ttv)*, which is required for the ability of Hh to reach target cells, encodes a glycosyltransferase. These enzymes synthesize glycosaminoglycan chains (GAGs) attached to the protein core of Heparan Sulfate Proteoglycans (HSPGs). The specificity of Ttv to Hh signaling suggests the existence of Hh-specific HS-GAGs. We propose that such HS-GAGs are synthesized by a complex that includes Ttv and a specific GAG-modifying enzyme(s) such as a sulfotransferase. The possible roles of HSPGs in Hh movement are discussed.

(a) Introduction

During development, a number of secreted factors have been identified that play critical roles in the patterning of field of cells. In particular, members of the Wnt, and Hedgehog (Hh) families have been shown to act long range to instruct cells of their specific fates. Despite extensive studies, the mechanisms underlying the movement of these molecules through tissues is not understood. Thus it is not clear whether the gradients of activities that these proteins trigger are established through diffusion of the secreted factors

in the extracellular space or through transport mechanisms that involve for example vesicle-like structures which are endocytosed and/or transcytosed.

In the case of Wnt proteins there is evidence that these molecules can be secreted in the extracellular space as well as transported through cells. Wnt proteins are poorly secreted in the extracellular space and bind the extracellular matrix tightly (Bradley and Brown, 1990; Gonzalez et al., 1991; Reichsman et al., 1996). These features suggest that these proteins may not be able to freely diffuse and thus raise the questions of how many cells away they are able to act. Studies in the *Drosophila* wing imaginal discs have provided compelling evidence that Wingless (Wg), the homolog of mammalian Wnt-1, is required directly for patterning a field of cells such as the wing blade. Further immunohistochemical analyses have detected Wg protein up to 25 cell diameters from the wing margin, which is the site of Wg synthesis in this tissue (Cadigan et al., 1998). Thus, Wg proteins can be found far away from *wg*-expressing cells raising the issue of the mechanism underlying the movement of Wg through cells. Interestingly, electron microscopy studies detect little Wg protein free in the extracellular space and Wg is present in vesicle like structures (Gonzalez et al., 1991). The presence of Wg in vesicles is endocytosis dependent and is not detected in *shibire* mutants which are defective in endocytosis (Bejsovec and Wieschaus, 1995). This observation has led to the model that Wg proteins are transported through cells in vesicle-like structures, a process referred to as transcytosis (Gonzalez et al., 1991).

Further support for this model come from the identification of several *wg* mutations which are defective in either Wg transport or signaling (Bejsovec and Wieschaus, 1995; Hays et al., 1997). Analysis of wg^{PE4} mutants has revealed that the Wg^{PE4} protein is competent to generate both short and long-range Wg effects when provided ectopically, suggesting that Wg^{PE4} is specifically defective in transport. Another mutant Wg protein, Wg^{CE7}, is distributed over many cell diameters on either side of *wg*-expressing cells, and is internalized into cells without transducing any of the known signaling activites of Wg. These results suggest that different Wg protein domains interact with different receptors; one involved in Wg transport and the other one in Wg signaling. Altogether, these studies have led to the model that the mechanism for long range movement of Wnt proteins relies on endocytosis and/or transcytosis.

Similarly, Hh proteins can travel long range. For example, in *Drosophila* Hh induces the expression of its target genes *patched (ptc)* and *decapentaplegic (dpp)* directly in a broad domain of 8-10 cell diameters along the anterior-posterior (A/P) boundary (Mullor et al., 1997; Strigini and Cohen, 1997). Further, Chen and Struhl (1996) (Chen and Struhl, 1996) have shown that in the absence of the Ptc receptor, Hh molecules are able to induce *ptc* transcription in wild type cells located beyond the clone of *ptc* mutant cells. Consistent with this result, we were able to detect co-immunolocalization of both Hh and Ptc in wild type cells located distally to the clone of *ptc* mutant cells (Bellaiche et al., 1998).

Hh is made as a precursor protein, which is autocatalytically cleaved to produce an N-terminal (Hh-N) and a C-terminal (Hh-C) fragment (Lee et al., 1994). During this cleavage process a cholesterol moeity is added to Hh-N (HhNp). This cholesterol modification anchors Hh to the membrane and prevents its secretion. HhNp is responsible for all the biological activities of Hh in flies and vertebrates. In the absence of the cholesterol modification, Hh-N molecules act at a longer range and are more potent inducers of Hh-target genes than HhNp (Porter et al., 1996). How HhNp can exert its patterning activity when it is associated to cholesterol, and thus does not have the properties of a diffusible molecule, is not clear. Futhermore, there are *in vitro* data that mammalian Hh is palmitoylated (Pepinsky et al., 1998). Such a modification, if it occurs *in vivo*, would also prevent secretion and free diffusion of Hh in the extracellular space. Roelink et al. (Roelink et al., 1995) have reported that, in tissue culture cells transfected with Sonic Hh (Shh), only very low concentrations of Shh can be detected in the medium. Thus, based on these observations it is not clear how Hh can act at long distance directly. One model is that Hh proteins are very potent signaling molecules requiring very low concentrations of diffusible Hh to regulate the activity of the Ptc receptor. According to this model some cholesterol-modified Hh molecules may be released from the plasma membrane, or that some mehanism exists that releases HhN from HhNp. Another model is that the cholesterol-modified form of Hh may be transported to target cells by a specific mechanism.

Biochemical analyses have indicated that Wnts and Hh are Heparin-binding proteins (Bradley and Brown, 1990) (Reichsman et al., 1996) (Bumcrot et al., 1995). While the function of binding to Heparin in the case of Hh molecules is unclear, in the case of Wnt proteins, there is now substantial biochemical and genetic evidence to support the

model that they interact with Heparan Sulfate Proteoglycans (HSPGs) *in vivo* (Binari et al., 1997; Cumberledge and Reichsman, 1997; Hacker et al., 1997; Haerry et al., 1997). HSPGs are composed of a protein core that is linked to glycosaminoglycan (GAGs) chains. These large macromolecules can be found at the cell surface, or in the extracellular matrix. HS can be covalently linked to a variety of cell surface proteins but is found consistently on two major families of proteoglycans, Syndecans and Glypicans (David, 1993) . The biological roles of HSPGs are highly diversified, ranging from simple mechanical support to as yet poorly understood effects on various cellular processes such as cell adhesion, motility, proliferation, differentiation and morphogenesis. In the context of signal transduction, HSPGs have been implicated in a number of events that include co-receptors for insoluble ligands, co-receptors for soluble ligands, internalization of receptors, transport of molecules or as soluble paracrine effectors (Salmivirta et al., 1996).

The assembly of HSPGs occurs in the golgi, when GAGs are added to specific serine residues on the protein core (Lindahl et al., 1998; Salmivirta et al., 1996). GAG chains consists of alternating sugar chains of D-glucuronic (GlcA) and N-acetyl-D-glucosamine (GlcNAc) units, which are joined together by a glucosyltransferase. The final GAG structures are formed by a series of polymer-modification reactions that are initiated by N-deacetylation and N-sulfation of GlcNAc units, followed by C-5 epimerization of D-glucuronic acid (GlcA) to L-iduronic acid (IdoA) and finally by incorporation of O-sulfate groups at various positions. The reactions involved in this process are generally incomplete such that a fraction of the potential substrate residues are not modified. Therefore, GAGs display considerable sequence heterogeneity both within and between different chains. The number of GAG chains on a protein core may also vary in length and number.

Recently, the critical roles of HSPGs in developmental processes has been illustrated by the identification of a number of mutations in enzymes involved in HSPGs biosynthesis. Mice lacking the HS 2-O-sulfotransferase exhibit a renal agenesis phenotype (Bullock et al., 1998). In *Drosophila*, mutations in *sugarless* (*sgl*, encoding UDP-glucose dehydrogenase) (Binari et al., 1997; Hacker et al., 1997; Haerry et al., 1997) and *pipe* (encoding a 2-O-sulfotransferase) (Sen et al., 1998) show defects in segmentation and establishment of dorsal/ventral (D/V) polarity in the embryo, respectively. While mutations in *pipe* only affect Toll signaling, in *sgl* mutants both Wg and FGF signaling pathways have been shown to be perturbed (Hacker et al., 1997; Lin et al., 1999).

Previously we have reported that the *tout velu* (*ttv*) *Drosophila* gene is required for the ability of Hh to reach target cells (Bellaiche et al., 1998). Hh was not able to spread through clones of cells that lack both *ptc* and *ttv* suggesting that Ttv activity is required for Hh movement. *ttv* is a member of the *Ext* gene family which have been implicated in the human multiple exostoses (Ext) syndrome(Ahn et al., 1995; Stickens et al., 1996). Recently, Ext proteins have been implicated in HSPGs biosynthesis. First, McCormick et al. (1998) (McCormick et al., 1998) have reported that the inability of the Herpes Simplex Virus to infect the mouse cell line sog9 correlates with a defect in HS biosynthesis. This defect, however, can be rescued by introducing an *Ext1* cDNA into the mutant cell line. Second, Lind et al. (1998) (Lind et al., 1998) have isolated an *Ext2* homolog as a glycosyltransferase required for the biosynthesis of HS from bovine serum. Altogether, these studies have led to the model that Ext proteins encode HS polymerase enzymes that joins the GlcA to GlcNAc saccharides together. Here we show *in vivo* evidence that *ttv* is involved in HS-GAG biosynthesis, thus providing the first evidence that HSPGs play a critical role in Hh signaling. Since other growth factors such as members of the Wnt and FGF families require HSPGs for signaling, we investigated the specificity of Ttv to Hh signaling. We demonstrate that *ttv* does not affect Wg and FGF signaling pathways suggesting that there is a specific binding sequence for Hh on HSPGs, and that Ttv is required to generate that sequence. We have identified another *Drosophila Ext* gene and, to account for the specificity of Ttv to Hh signaling, we speculate that Hh specific HS-GAGs are synthesized by a complex that includes Ttv and at least one specific sulfotransferase.

(b) Results

Tout velu is involved in HSPGs biosynthesis

ttv is 56% homologous to *Ext1*, and 26% homologous to *Ext2*, *ttv* is a zygotic lethal mutation and animals derived from heterozygous *ttv* females die at the pupal stage. However, when both maternal and zygotic *ttv* activities are removed following the generation of females with *ttv* germline clones (see Experimental Procedures), *ttv* embryos die during embryogenesis with a segment polarity phenotype (see below). We refer to these embryos as "*ttv* embryos" in the text. We have previously shown that in *ttv* mutant clones in the wing imaginal disc Hh movement is affected (Bellaiche et al., 1998). Recently, two reports have implicated Ext proteins in HSPGs biosynthesis (McCormick et al., 1998, Lind, 1998 #1). Since both reports have only shown indirect biochemical evidence that

Exts are glycosyltransferases, we investigated the *in vivo* function of the *Drosophila* Ext homolog *ttv*. Therefore, we analyzed whether HS biosynthesis was affected in the absence of *ttv* activity.

To detect the presence of HSPGs *in vivo*, we stained embryos with an antibody (3G10) which recognizes an epitope generated following digestion of HSPG with heparinase III. Following this digestion of the HS sugar chains, one desaturated uronate residues per chain will remain linked to the core protein, and therefore the 3G10 antibody will trace the core protein (David et al., 1992). Staining of wild type embryos with 3G10 reveals a uniform staining during early stages and a more pronounced central nervous system staining at later stages (Figs 6B and C). No staining could be detected in embryos which have not been treated with heparinase III (Figs 6A and D). In contrast to wild type, the staining detected by 3G10 in *ttv* embryos was strongly reduced (Fig. 6E). However, this staining was recovered when wild type *ttv* activity is reintroduced zygotically in *ttv* embryos, as shown for example in Fig. 6F where *ttv* expression is under the control of the promoter of the pair rule gene *hairy* (*h-ttv*). Finally, using this assay, we found that *ttv* is not a rate limiting enzyme in HS biosynthesis because no additional staining was found within the *hairy* stripes when *h-ttv* is expressed in wild type embryos,

The reduced staining detected by the 3G10 monoclonal antibody in *ttv* embryos (Fig 6E) could be due to residual Ttv activity. Therefore we determined whether the *ttv* $^{l(2)06681}$ allele used in our experiments is a complete loss of function mutation. We crossed females with *ttv* germline clones to males heterozygous for *Df(2R)Trix*, a deficiency that uncovers the *ttv* locus. The segment polarity cuticle phenotype (see below) of *ttv*$^{l(2)00681}$ /*Df(2R)Trix* embryos is similar to the phenotype of *ttv*$^{l(2)00681}$ homozygous embryos (Fig. 6H and I) indicating that the *ttv*$^{l(2)06681}$ allele behaves as a genetic null. In addition, we did not detect any Ttv protein in Western blots prepared from *ttv*$^{l(2)00681}$ embryos (Fig. 6J), indicating that the *ttv*$^{l(2)06681}$ allele behaves as a protein null allele.

Because *ttv*$^{l(2)06681}$ behaves as a complete loss of function allele, an explanation for the 3G10 reduced staining in *ttv* embryos is that there might be at least one other *Ext*-like gene in the *Drosophila* genome. We found one *Drosophila* EST database with some homology to *ttv*. When the cDNA clone was sequenced, we found that it was more

homologous to vertebrate Ext2 (44% protein identity) than to Ext1 (26% protein identity) (Fig. 7). Therefore we have named the gene corresponding to this cDNA *DExt2*. Since *DExt2* and *ttv* are maternally expressed and uniformly distributed in early embryos (not shown), we propose that the residual 3G10 staining present in *ttv* embryos is due to activity of this second *Ext*-like gene. In conclusion, our data indicate that there are at least two Ext proteins, Ttv/DExt1 and DExt2, expressed in *Drosophila* embryos. In the complete absence of Ttv/DExt1, HSPGs biosynthesis, as detected by the 3G10 antibody, is reduced.

Subcellular localization of Ttv

Ttv encodes a type II protein (Bellaiche et al., 1998) which could be localized at the plasma membrane or in the membranes of the secretory apparatus. To determine the cellular localization of Ttv, we generated a polyclonal antibody against Ttv (see Experimental Procedures). Although the antibody recognizes Ttv on Western Blots (Fig. 6H), we could not obtain a convincing staining pattern in embryos or in imaginal discs. To circumvent this problem, we tagged the Ttv protein with 6 myc-epitopes at the C-terminus of the *ttv* coding sequence (*ttvmyc*). This *ttvmyc* gene when expressed under the control of the UAS promoter using the Gal4-UAS system (Brand and Perrimon, 1993) is able to rescue the segment polarity phenotype of *ttv* embryos (data not shown).

Staining of embryos or imaginal discs expressing the *UAS-ttvmyc* construct in different patterns with antibodies against the Myc-epitope did not reveal any plasma membrane staining; however, perinuclear and punctated staining was obvious (Fig 8A). To address in which subcellular compartments the Ttv protein can be found, we tested whether Ttv-myc colocalizes with either the Golgi protein ß -copII (Fig. 8B), or the Endoplasmic Reticulum (ER) protein Bip (Fig. 8C). Interestingly, Ttv-myc mostly colocalizes with Bip and partially with ß-CopII (Fig.38 D). Further, we did not detect a colocalization with the plasma membrane markers E-cadherin or Armadillo (not shown). Altogether, these results indicate that Ttv resides mainly in the ER and in the Golgi, which is consistent with its function as an enzyme involved in HSPGs biosynthesis.

Finally, because we previously demonstrated that Ttv affects Hh signaling, we performed co-immunolocalization studies between Ttv-myc and Hh, as well as between Ttv-myc and Ptc. In these experiments, we could detect the Ttv-myc protein at the apical, but also basolateral compartments of cells, and the Ttv-myc protein did not colocalize with either Hh (Fig. 8E and F) or Ptc (Fig. 8G and H). Thus, because Ttv affects HS

biosynthesis, we conclude that its effect on Hh diffusion is indirect and mediated through the interaction between Hh and a HSPGs.

FGF and Wg signaling is not affected in ttv embryos

HSPGs have been implicated in the signaling pathways of several growth factors. In *Drosophila*, *sgl* and *sulfateless* (*sfl*) mutations that affect HSPGs biosynthesis affect both the Wg and FGF pathways (Hacker et al., 1997; Lin et al., 1999). Therefore, to further address the specificity of Ttv to Hh signaling, we examined in details the effect of loss of *ttv* activity on both FGF and Wg signaling pathways.

To investigate if FGF signaling is affected in *ttv* embryos, we examined the migration of the mesoderm, a process dependent on the FGF receptor Heartless (Htl) signaling pathway (Gisselbrecht et al., 1996). In wild type embryos, the ventral mesoderm invaginates at stage 6 and at stage 9 mesoderm cells rearrange and form a monolayer. Mutants with defects in Htl signaling show aberrant mesodermal cell migration and mesodermal cells do not form a monolayer. Similarly, mutations in *sgl* and *sfl* also exhibit a mesoderm migration defect consistent with the role of HSPGs in FGF/Htl signaling (Lin et al., 1999). However, surprisingly for a mutation involved in HS biosynthesis, we could not detect any defects in mesoderm migration in *ttv* embryos (Fig. 9A-D). Migration of Twist (Twi) positive mesodermal cells is normal as well as the domain of Twi staining which indicates that there is no D/V patterning defect in *ttv* embryos.

Because HSPGs have been implicated in Wg signaling (Binari et al., 1997; Hacker et al., 1997; Haerry et al., 1997), we examined whether Wg signaling is affected in the absence of Ttv activity. In the embryo, Wg signaling is involved in a number of developmental processes that include the maintenance of En expression. *ttv* embryos display a segment polarity phenotype similar to loss of Wg signaling, as shown by both the absence of naked cuticle (Fig 6H) and disapearance of both *wg* and *en* expression (Fig. 10A-D). However, since Wg and Hh signaling pathways in the embryonic epidermis are dependent on each others, the segment polarity phenotype does not allow us to distinguish whether loss of Ttv affects either Wg or/and Hh signaling. Therefore, we looked at other Wg-dependent processes during embryogenesis, like the formation of the SNS and formation of the RP2 neurons. In wild type embryos, the invagination of the three SNS neurons can be visualized by staining with antibodies against Crumbs. In mutants that decrease Wg signaling there is less than three SNS invaginations, while in mutants which

increase Wg signaling more than three invaginations can be found (Gonzalez-Gaitan and Jackle, 1995). The SNS phenotype of *ttv* mutant embryos appear wild type (Fig 10E-G), suggesting that Wg signaling is not affected by loss of Ttv activity. Similarly, we could not detect a requirement for Ttv activity in formation of the RP2 neurons. In *ttv* embryos, the correct number of RP2 neurons, detected using an Even-skipped (Eve) antibody, are found (not shown). Thus, several Wg dependent processes are not affected in *ttv* embryos.

To extend the conclusion that Ttv is not required for Wg, we examined the effect of loss of Ttv function in Wg signaling at another developmental stage: the wing imaginal discs. In the wing, Wg expressed at the wing margin controls patterning along the D/V axis in a concentration dependent manner. Expression of the proneural *Achaete -scute* gene at the margin, and *vestigial (vg)* and *distalless (dll)* at a longer distance are controlled by Wg (Neumann and Cohen, 1997; Zecca et al., 1995). In *ttv* mutant clones the expression of A101, a proneural marker induced by Wg in the wing margin is not affected (not shown). Further, we find that the expression of *dll* is not affected (Fig. 10H and I). Thus, as concluded from the results in the embryo, Wg signaling in wing imaginal discs does not require Ttv activity.

Ttv is required for Hh diffusion

The expression pattern of the gene *bagpipe* (*bap)* also provides evidence that Hh signaling and not Wg is affected in *ttv* embryos. The *bap* segmental expression at stage 10 during the formation of the mesoderm is dependent upon both Wg and Hh signaling. Wg is required for repression of *bap* expression between the segments, and in *wg* mutant embryos *bap* expression becomes continuous. Hh is required for maintenance of *bap* expression and in *hh* mutant embryos *bap* expression fades. In *ttv* embryos, *bap* staining disappears providing further evidence that *ttv*is involved in Hh and not Wg signaling in the embryo (Fig. 11A-D).

Previously, we have documented that Ttv is required for the ability of Hh to reach target cells (Bellaiche et al., 1998). To extend these observations to embryonic stages, we examined Hh expression in *ttv* embryos. Staining wild type embryos with a Hh antibody shows a strong staining in Hh-expressing cells and a punctate staining outside of these cells (Figure 11E and F). However, in *ttv* embryos, Hh is only seen in Hh producing cells. This effect appears specific to Hh because Wg diffusion is not impaired in *ttv* embryos (Fig. 10J and K).

Hh is produced as a precursor protein, which undergoes autocleavage (see Introduction). During this process a cholesterol moeity is attached to the N-terminal part of Hh (HhNp) which encodes the signaling domain of Hh. Since HhNp has a cholesterol anchor, it is presumed to remain bound to the membrane. Porter et al. (Porter et al., 1995) have shown that HhN can move further in embryos than HhNp, and induce ectopic *wg*-expression. To determine whether the effect of *ttv* on Hh diffusion depends on the cholesterol modification of Hh, we tested whether the diffusion of HhN is affected in *ttv* embryos. We expressed UAS-HhN under the control of *en-Gal4*. Interestingly, HhN diffuses (Fig. 11G) and induces ectopic *en* expression (Fig. 11H). Thus, *ttv* is required for the proper diffusion of the cholesterol modified HhNp but not of HhN.

(c) Discussion

In this exmaple, we report that Ttv activity is required for the formation of HSPGs in *Drosophila*, an activity consistent with the proposal that Ext proteins encode glycosyltransferase enzymes. We show that, as expected for such an enzyme, a tagged form of Ttv which is associated with biological activity, localizes mainly in the ER and the Golgi. Our phenotypic analysis of *ttv* mutants reveals that Hh signaling is pertubed, providing the first direct evidence that HSPGs are required for proper signaling of the heparin-binding protein Hh. Further, because only Hh, but not other signaling pathways previously shown to require HSPGs, is affected in *ttv* mutants, our results suggest that *ttv* is required to synthesize a specific GAG binding sequence for Hh on HSPGs.

Ttv is specific to Hh signaling

Using the 3G10 antibody, we have shown that in the absence of *ttv* activity the amount of HSPGs is reduced. This result is consistent with the proposed function of the mammalian Ext proteins in HSPGs biosynthesis. Interestingly, in *ttv* mutant embryos not all HSPGs are absent because there is residual staining detected by 3G10. This observation suggests that at least one other *Drosophila* glycosyltransferase is expressed in embryos, an observation that we have substantiated by the characterization of *DExt2*.

Although a bovine Ext2 protein has been isolated as a GlcA- and GlcNAc transferase, there is evidence that the activity for the GlcA and GlcNAc transferase is mediated by two different enzymes (Lidholt and Lindahl, 1992). However, a single mutational event was found to eliminate both transferase activities in tissue culture cells,

suggesting that the two enzymes might share a common subunit (Lidholt et al., 1992). Therefore it is not entirely clear what function Ext proteins have in HSPGs biosynthesis.

Previously, we reported that in the absence of either *sgl* (which encodes UDP-glucose dehydrogenase) or *sfl* (which encodes a N-sulfotransferase) both Wg and FGF signaling pathways are affected. Our analysis of *ttv* suggests that Hh signaling should be affected as well in these mutants and indeed we have recently found this to be the case (Lin et al., in preparation). Because these enzymes are required for the formation of the HS-GAG chains present on HSPGs, we expected both Wg and FGF signaling to be affected as well in *ttv* mutants. However, our detailed phenotypic analyses of *ttv* mutant revealed that only Hh signaling is perturbed in the absence of Ttv activity. This result cannot be attributed to differential sensitivity of these pathways to a reduced amount of HS-GAGs synthesized by Ttv because the $ttv^{l(2)00681}$ allele we used in most of our analyses is a genetic null and no Ttv protein can be detected in Western blots prepared from $ttv^{l(2)00681}$ mutant animals.

GAGs chain specificity

Because of the existence of at least another *Drosophila Ext* gene, the observation that Hh signaling but not FGF and Wg signaling is affected in the absence of Ttv activity, could indicate that Hh signaling is more sensitive to a reduction of HS-GAGs than Wg and FGF signaling. According to this "quantitative" model, Wg and FGF signaling pathways would not be affected in *ttv* embryos because HS-GAGs synthesized by another *Drosophila* Ext are sufficient to allow these pathways to function. Alternatively, according to a "qualitative" model, the specificity of Ttv to Hh signaling suggests the existence of Hh-specific HS-GAGs. In *ttv* mutants Wg and FGF signaling may not be affected because the HS-GAGs that these factors bind to are present. We favor the "qualitative" model; i.e., that Ttv activity is required for the synthesis of an Hh-specific HS-GAGs, because according to the quantitative model we would still expect Wg and FGF signaling pathways to be at the least partially affected. However, this is not the case as we found no evidence that the activity of these pathways is affected. Further, in the absence of Ttv activity, the effect on Hh signaling is similar to the loss of Hh activity.

All HS-GAGs consist of an alternative chain of GlcA and GlcNAc residues which are linked together by the glycosyltransferase/Ext enzymes. A puzzle is why there are

several *Ext*-like genes in *Drosophila* and vertebrates if their functions are completely redundant. This diversity in *Ext* genes does not reflect the tissue specific expression pattern of the *Ext* genes because both *ttv* and *DExt2* are uniformly expressed during embryogenesis as assayed by RNA *in situ*. To account for the specificity of Ttv to Hh signaling as well as the existence of multiple *Ext* genes, we propose that different Ext glycosyltransferases form specific complexes in the golgi with different GAG modifying enzymes. Each complex might generate qualitatively different GAG sequences which will specifically bind certain growth factors. Because all Ext proteins may have the same biochemical activities, we propose that the specificity of the GAG sequences is generated by specific sulfotransferases. Interactions between HS and proteins depend on the presence of sulfate groups. Binding of the different members of the FGF family, for example, may require different combinations of sulfate groups, hence different saccharide sequences. Such sequences may be represented in the same GAG chains, they may also be differentially expressed in separate GAG chains of the HS. The generation of such specific sugar chains would require selective polymer modification. Thus, Ttv might form a complex with enzymes which generate a GAG sequence required for Hh binding only,which would explain why only Hh signaling is affected in *ttv* embryos. In support of this "enzyme complex" model, it has been proposed that chain elongation and modification occur simultaneously (Lidholt and Lindahl, 1992). A prediction of this model would be that there is a specific sulfotransferase which is associated in a complex with Ttv, to generate Hh specific binding sites on GAGs.

Further, there is now evidence that some sulfotransferases are associated with tissue specific functions. For example, in the *Drosophila* Toll pathway, *pipe*, which encodes for a 2-O-sulfotransferase is expressed in the ventral part of the egg chamber and has been proposed to activate the protease cascade that leads to production of the active Toll ligand Spätzle (Sen et al., 1998). The effect of *pipe* on Toll signaling is most likely indirect through a GAG attached to a proteoglycan. The role of the proteoglycan could either be activation or assembly of a serine protease complex which processes Spätzle or concentration of Spätzle to the ventral side of the egg chamber. Finally, mice lacking a HS 2-O-sulfotransferase undergo normal embryogenesis, but die around birth. They exhibit a fully penetrant defect in kidney development. In addition, the mice also have defects in the eye, skeleton, a cleft palate and polydactily of the limb (Bullock et al., 1998).

Role of HSPGs involved in Hh signaling

Our results indicate that HSPGs are involved in the ability of Hh to reach target cells. Although there is only a very low concentration of Hh detected outside Hh-producing cells, Hh can have its effect a few cell diameters wide. Therefore, either the concentration of Hh required to signal is very low and the low amount of diffusible Hh is sufficient for signaling or the membrane tethered Hh can be transported from cells to cells. One model in which HSPGs could influence Hh distribution is by concentrating Hh and perhaps presenting it to its receptor (Fig 12B1). Such a function has been proposed for HSPGs in FGF signaling. However, if HSPGs are simply concentrating Hh, one might expect that Hh could travel over a *ptc ttv* double mutant clone. This model assumes that HSPGs are not playing a more active role in the extracellular spreading of Hh.

Another possibility is that there is a transport mechanism for Hh which would allow Hh to move from cell to cell (Fig 12B2). The transport of Hh might involve so called rafts, which are microdomains in the plasma membrane rich in sphingolipids, cholesterol and GPI-anchored proteins (Simons and Ikonen, 1997). Since signaling molecules and their receptors have been shown to cluster in these rafts, it has been proposed that signaling might occur in these microdomains. Interestingly, Hh has been reported to localize into the raft fraction after separation of cell extracts (Rietveld et al., 1999). Perhaps a GPI-anchored HSPGs is required to localize Hh in these rafts. Transfer of GPI anchored proteins between cells have been observed and Hh might be transferred from cell to cell in this way (Kooyman et al., 1995). The cholesterol modification on Hh might also facilitate Hh localization into the rafts after which transport of Hh can occur.

Another model is that HSPGs play a role in a process which releases Hh from its cholesterol anchor and generates a non-membrane bound Hh (Fig 12B3). Since the HS 2-O-sulfotransferase encoded by *pipe* affects processing of Spätzle it is conceivable that HSPGs have a similar role in Hh processing. Our previous data in the wing disc, however, argues against a role for *ttv* in Hh processing. The processing of Hh presumably takes place in the Hh-producing cell. However, removing *ttv* in the Hh producing cells in the posterior compartment did not abolish Hh signaling in the anterior compartment (Bellaiche et al., 1998). This would also argue against a model in which a proteoglycan to which Hh is attached is cleaved and then shed. This model for Hh movement seems unlikely because *ttv* is cell autonomous and it is not clear how Hh could be released from the membrane.

Regardless of the mechanism of Hh movement, our results demonstrate that HSPGs are involved in the regulation of Hh distribution. It will be of interest to identify the protoglycan which facilitates Hh movement. Interestingly, there are several GPI-anchored proteoglycans in *Drosophila*. Loss of function mutations in one of them, Dally, seems to affect Dpp signaling (Penton et al., 1997). Interestingly, the first phenotype described for Dally, division abnormally delayed in the proliferation centers of the brain, is a phenotype which could be attributed to Hh as well (Nakato et al., 1995). Hh has been shown to be transported through the retinal axons and trigger proliferation in the brain (Huang and Kunes, 1996). Thus, it would be interesting to see if transport of Hh is affected in *dally* mutants.

Extracellular matrix, bone development and disease

Several growth factors like FGFs, BMPs, and also Indian Hh are important for skeletal growth regulation (Erlebacher et al., 1995; Lanske et al., 1996; Vortkamp et al., 1996). Mutations in these growth factors or their receptors have been linked to several bone malformity diseases. These growth factors are secreted into the extracellular matrix, where they diffuse to their target cells. One of the major component of extracellular matrix are GAGs and PGs. In bone, the macromolecules of the extracellular matrix in cartilage is secreted by chondroblasts and by osteoblasts and the matrix can become calcified to form a hard structure. Therefore, it is not surprising that mutations that affect proteoglycan biosynthesis can lead to skeletal disorders, like the multiple exostoses syndrome (EXT). *Ext* is probably not the only HSPG biosynthesis gene affecting bone morphognesis. Mice lacking 2-O sulfotransferase alxo exhibit skeletal abnormalities (Bullock et al., 1998). Interestingly, the Diastrophic Dysplasia (DTD) gene encodes a sulfate transporter and staining for proteoglycans is reduced in DTD patients [Hästbacka et al., 1994] . It is possible that other human bone disorders may result from mutations affecting other steps of proteoglycan biosynthesis.

(d) Experimental Procedures

Genetics

ttv mutant animals derived from heterozygous *ttv* females die at the pupal stage. However, when both maternal and zygotic *ttv* activities are removed, following the generation of females with *ttv* germline clones, *ttv* embryos die during embryogenesis with

a segment polarity phenotype. We refer to these embryos as "*ttv* embryos" in the text. The *ttv* maternal effect phenotype is paternally rescuable (Perrimon et al., 1996). Paternally rescued *ttv* embryos that have received a wild type copy of *ttv* from their fathers develop to adults.

Females with germline clones were generated using the "FLP-DFS" technique (Chou and Perrimon, 1996). Virgin females of the genotype FRT^{G13} *ttv/CyO* were mated with males of the genotype y w FLP^{12}/+, FRT^{G13} $P[ovo^{D1}]$/*CyO*. The resulting progeny was heat shocked at 37°C for 2 hrs during larval stages of development and *y w* FLP^{12}/+; FRT^{G13} *ttv* /FRT^{G13} $P[ovo^{D1}]$ females which carry homozygous *ttv* germline clones were selected. These females were mated to *ttv/CyO, ftz-LacZ* males and maternal/zygotic null embryos were identified by the absence of beta-galactosidase expression.

For expression of *ttv* in the *hairy* domain, *y w* FLP^{12}/+; FRT^{G13} *ttv* /FRT^{G13} $P[ovo^{D1}]$; *hairy-GAL4/+* females were crossed to *ttv/CyO, ftz-LacZ; UAS-ttv* males. For expressing HhN in *ttv* embryos, *y w* FLP^{12}/+; FRT^{G13} *ttv* /FRT^{G13} $P[ovo^{D1}]$; *UAS-HhN/+* females were crossed to *ttv en-GAL4/CyO, ftz-LacZ* males.

UAS constructs and GAL4 lines

The *UAS-ttv* construct is an SpeI/NotI fragment of the *ttv* gene which consist of the whole gene cloned into the XXX site of pUAST (Brand and Perrimon, 1993).The UAS-*ttv*myc construct is the same construct with 6 myc epitope tags at its C-terminus.

For ectopic expression the GAL4 insertions in hairy-GAL4 (1J3) on the third chromosome, en-GAL4 on the second chromosome in the endogenous genes were used.

In situ hybridization.

Antisense *ttv, DExt2* and *bap* probes were generated by the DIG labeling kit from Boehringer Mannheim using wither T3 or T7 RNA polymearse. RNA in situ hybridization was performed as described in (Tautz and Pfeifle 1989)

Detection of Ttv

For Western blot analyses, embryos were dechorionated with 50% bleach, dounced in 20 mM Tris HCL PH 7.5, 150 mM NaCl, 5mM EDTA, 1% Triton X-100 at 4 C with

protease inhibitors and centrifuged for 20 min at 14 000xg at 4°C. Amount of protein was measured using Biorads protein assay (Biorad) and 100 μg of protein was loaded in each lane, immunoblotting was performed as described in Harlow and Lane.

Immunohistochemistry

Embryonic cuticles were prepared by the Hoyer's mountant method (van der Meer, 1977). Fixation of embryos, antibody staining were performed as described in Patel, 1994. Treatment of embryos to expose the epitope recognized by the 3G10 monoclonal antibody (Seikagaku) was done with 500 mU/ml heparinaseIII (Sigma) in 50 mM Tris-HCL PH 7.2, 100 mM NaCl, 1mM Ca Cl2, 0.1% Triton-X100, 5 μg/ml BSA. Embryos were incubated for 16 hrs at 37° C and stained with 3G10 antibody. Secondary antibodies for histochemical staining were from Vectorlabs, fluorescent secondary antibodies from Jackson Immunoresearch. Embryos sections were performed as described in Gisselbrecht, 1996.

(e) References for example 5

Ahn, J., Ludecke, H. J., Lindow, S., Horton, W. A., Lee, B., Wagner, M. J., Horsthemke, B., and Wells, D. E. (1995). Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). Nat Genet *11*, 137-143.

Bejsovec, A., and Wieschaus, E. (1995). Signaling activities of the Drosophila wingless gene are separately mutable and appear to be transduced at the cell surface. Genetics *139*, 309-20.

Bellaiche, Y., The, I., and Perrimon, N. (1998). Tout-velu is a Drosophila homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion [see comments]. Nature *394*, 85-8.

Binari, R. C., Staveley, B. E., Johnson, W. A., Godavarti, R., Sasisekharan, R., and Manoukian, A. S. (1997). Genetic evidence that heparin-like glycosaminoglycans are involved in wingless signaling. Development *124*, 2623-32.

Bradley, R. S., and Brown, A. M. (1990). The proto-oncogene int-1 encodes a secreted protein associated with the extracellular matrix. Embo J *9*, 1569-75.

Brand, A. H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-15.

Bullock, S. L., Fletcher, J. M., Beddington, R. S., and Wilson, V. A. (1998). Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase. Genes Dev *12*, 1894-906.

Bumcrot, D. A., Takada, R., and McMahon, A. P. (1995). Proteolytic processing yields two secreted forms of sonic hedgehog [published erratum appears in Mol Cell Biol 1995 May;15(5):2904]. Mol Cell Biol *15*, 2294-303.

Cadigan, K. M., Fish, M. P., Rulifson, E. J., and Nusse, R. (1998). Wingless repression of Drosophila frizzled 2 expression shapes the Wingless morphogen gradient in the wing. Cell *93*, 767-77.

Chen, Y., and Struhl, G. (1996). Dual roles for patched in sequestering and transducing Hedgehog. Cell *87*, 553-63.

Cumberledge, S., and Reichsman, F. (1997). Glycosaminoglycans and WNTs: just a spoonful of sugar helps the signal go down. Trends Genet *13*, 421-3.

David, G. (1993). Integral membrane heparan sulfate proteoglycans. Faseb J *7*, 1023-30.

David, G., Bai, X. M., Van der Schueren, B., Cassiman, J. J., and Van den Berghe, H. (1992). Developmental changes in heparan sulfate expression: in situ detection with mAbs. J Cell Biol *119*, 961-75.

Erlebacher, A., Filvaroff, E. H., Gitelman, S. E., and Derynck, R. (1995). Toward a molecular understanding of skeletal development [comment]. Cell *80*, 371-8.

Gisselbrecht, S., Skeath, J. B., Doe, C. Q., and Michelson, A. M. (1996). heartless encodes a fibroblast growth factor receptor (DFR1/DFGF-R2) involved in the directional migration of early mesodermal cells in the Drosophila embryo. Genes Dev *10*, 3003-17.

Gonzalez, F., Swales, L., Bejsovec, A., Skaer, H., and Martinez Arias, A. (1991). Secretion and movement of wingless protein in the epidermis of the Drosophila embryo. Mech Dev *35*, 43-54.

Gonzalez-Gaitan, M., and Jackle, H. (1995). Invagination centers within the Drosophila stomatogastric nervous system anlage are positioned by Notch-mediated signaling which is spatially controlled through wingless. Development *121*, 2313-25.

Hacker, U., Lin, X., and Perrimon, N. (1997). The Drosophila sugarless gene modulates Wingless signaling and encodes an enzyme involved in polysaccharide biosynthesis. Development *124*, 3565-73.

Haerry, T. E., Heslip, T. R., Marsh, J. L., and O'Connor, M. B. (1997). Defects in glucuronate biosynthesis disrupt Wingless signaling in Drosophila. Development *124*, 3055-64.

Hays, R., Gibori, G. B., and Bejsovec, A. (1997). Wingless signaling generates pattern through two distinct mechanisms. Development *124*, 3727-36.

Huang, Z., and Kunes, S. (1996). Hedgehog, transmitted along retinal axons, triggers neurogenesis in the developing visual centers of the Drosophila brain. Cell *86*, 411-22.

Kooyman, D. L., Byrne, G. W., McClellan, S., Nielsen, D., Tone, M., Waldmann, H., Coffman, T. M., McCurry, K. R., Platt, J. L., and Logan, J. S. (1995). In vivo transfer of GPI-linked complement restriction factors from erythrocytes to the endothelium. Science *269*, 89-92.

Lanske, B., Karaplis, A. C., Lee, K., Luz, A., Vortkamp, A., Pirro, A., Karperien, M., Defize, L. H. K., Ho, C., Mulligan, R. C., Abou-Samra, A. B., Juppner, H., Segre, G. V., and Kronenberg, H. M. (1996). PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth [see comments]. Science *273*, 663-6.

Lee, J. J., Ekker, S. C., von Kessler, D. P., Porter, J. A., Sun, B. I., and Beachy, P. A. (1994). Autoproteolysis in hedgehog protein biogenesis [see comments]. Science *266*, 1528-37.

Lidholt, K., and Lindahl, U. (1992). Biosynthesis of heparin. The D-glucuronosyl- and N-acetyl-D- glucosaminyltransferase reactions and their relation to polymer modification. Biochem J *287*, 21-9.

Lidholt, K., Weinke, J. L., Kiser, C. S., Lugemwa, F. N., Bame, K. J., Cheifetz, S., Massague, J., Lindahl, U., and Esko, J. D. (1992). A single mutation affects both N-acetylglucosaminyltransferase and glucuronosyltransferase activities in a Chinese hamster ovary cell mutant defective in heparan sulfate biosynthesis. Proc Natl Acad Sci U S A *89*, 2267-71.

Lin, X. H., Buff, E. M., Perrimon, N. P., and Michelson, A. M. (1999). Heparan sulfate proteoglycans are essential for FGF receptor signaling during Drosophila embryonic development. submitted.

Lind, T., Tufaro, F., McCormick, C., Lindahl, U., and Lidholt, K. (1998). The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. J Biol Chem *273*, 26265-8.

Lindahl, U., Kusche-Gullberg, M., and Kjellen, L. (1998). Regulated diversity of heparan sulfate. J Biol Chem *273*, 24979-82.

McCormick, C., Leduc, Y., Martindale, D., Mattison, K., Esford, L. E., Dyer, A. P., and Tufaro, F. (1998). The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate [see comments]. Nat Genet *19*, 158-61.

Mullor, J. L., Calleja, M., Capdevila, J., and Guerrero, I. (1997). Hedgehog activity, independent of decapentaplegic, participates in wing disc patterning. Development *124*, 1227-37.

Nakato, H., Futch, T. A., and Selleck, S. B. (1995). The division abnormally delayed (dally) gene: a putative integral membrane proteoglycan required for cell division patterning during postembryonic development of the nervous system in Drosophila. Development *121*, 3687-702.

Neumann, C. J., and Cohen, S. M. (1997). Long-range action of Wingless organizes the dorsal-ventral axis of the Drosophila wing. Development *124*, 871-80.

Penton, A., Selleck, S. B., and Hoffmann, F. M. (1997). Regulation of cell cycle synchronization by decapentaplegic during Drosophila eye development. Science *275*, 203-6.

Pepinsky, R. B., Zeng, C., Wen, D., Rayhorn, P., Baker, D. P., Williams, K. P., Bixler, S. A., Ambrose, C. M., Garber, E. A., Miatkowski, K., Taylor, F. R., Wang, E. A., and Galdes, A. (1998). Identification of a palmitic acid-modified form of human Sonic hedgehog. J Biol Chem *273*, 14037-45.

Porter, J. A., Ekker, S. C., Park, W. J., von Kessler, D. P., Young, K. E., Chen, C. H., Ma, Y., Woods, A. S., Cotter, R. J., Koonin, E. V., and Beachy, P. A. (1996). Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. Cell *86*, 21-34.

Porter, J. A., von Kessler, D. P., Ekker, S. C., Young, K. E., Lee, J. J., Moses, K., and Beachy, P. A. (1995). The product of hedgehog autoproteolytic cleavage active in local and long-range signalling. Nature *374*, 363-6.

Reichsman, F., Smith, L., and Cumberledge, S. (1996). Glycosaminoglycans can modulate extracellular localization of the wingless protein and promote signal transduction. J Cell Biol *135*, 819-27.

Rietveld, A., Neutz, S., Simons, K., and Eaton, S. (1999). Association of the cholesterol-linked N-terminal fragment of Hedgehog domain with Drosophila raft lipid microdomain. J. Biol. Chem. *in press*.

Roelink, H., Porter, J. A., Chiang, C., Tanabe, Y., Chang, D. T., Beachy, P. A., and Jessell, T. M. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. Cell *81*, 445-55.

Salmivirta, M., Lidholt, K., and Lindahl, U. (1996). Heparan sulfate: a piece of information. Faseb J *10*, 1270-9.

Sen, J., Goltz, J. S., Stevens, L., and Stein, D. (1998). Spatially restricted expression of pipe in the Drosophila egg chamber defines embryonic dorsal-ventral polarity. Cell *95*, 471-81.

Simons, K., and Ikonen, E. (1997). Functional rafts in cell membranes. Nature *387*, 569-72.

Stickens, D., Clines, G., Burbee, D., Ramos, P., Thomas, S., Hogue, D., Hecht, J. T., Lovett, M., and Evans, G. A. (1996). The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. Nat Genet *14*, 25-32.

Strigini, M., and Cohen, S. M. (1997). A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. Development *124*, 4697-705.

Vortkamp, A., Lee, K., Lanske, B., Segre, G. V., Kronenberg, H. M., and Tabin, C. J. (1996). Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein [see comments]. Science *273*, 613-22.

Zecca, M., Basler, K., and Struhl, G. (1995). Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the Drosophila wing. Development *121*, 2265-78.

Example 6: Dally, a member of the glypican family of Heparan sulfate proteoglycans, regulates Wingless signaling in *Drosophila*

Characterization of the *Drosophila* gene *sulfateless*, which encodes an homolog of vertebrate heparan sulfate N-deacetylase/N-sulfotransferase - an enzyme essential for the modification of heparan sulfate (HS), reveals that HS proteoglycans (HSPGs) are necessary for Wg signaling. We have identified Dally, a GPI-linked Glypican, as the HSPG molecule involved in Wg signaling. Loss of *dally* activity, both in the embryo and imaginal dics, generates phenotypes reminiscent to loss of Wg activity. Interestingly, *dally* is co-

expressed with the Wg receptor Dfz2. We propose that Dally serves as a co-receptor for Dfz2, and together with Dfz2 modulates both short and long-range activities of Wg.

(a) Introduction

During the development of multicellular organisms, the generation of complex patterns is controlled by several key signaling molecules including members of the Wnt, Hedgehog (Hh), TGFb, TGFalpha and FGF families. The secretion, distribution and subsequent signaling of these signaling molecules on the cell surface play essential roles for the coordination of cell growth and differentiation. These signaling molecules usually bind tightly to the cell surface upon secretion and some of them can serve as morphogens triggering intracellular events in a concentration-dependent manner to specify cell fates. While the intracellular events regulated by these signaling molecules have been extensively characterized, it is not well understood how they are regulated from the time of their synthesis to forming active complexes with their receptors.

One of the major class of cell surface molecules are heparan sulfate proteoglycans (HSPGs). HSPGs are ubiquitous macromolecules associated with both the cell surface and the extracellular matrix (ECM). HSPGs consist of a protein core to which Heparin/HS glycosaminoglycans (HS GAGs) are attached (reviewed by Bernfield et al., 1992; David, 1993; Kjellen and Lindahl, 1991; Yanagishita and Hascall, 1992). The glycosyl phosphatidylinositol (GPI) linked Glypicans and the transmembrane Syndecans represent the two major cell surface HSPGs. While Syndecans bear both HS and chondroitin sulfate (CS) GAGs, Glypicans are exclusively attached to HS GAGs (Bernfield et al., 1992; David, 1993). Through a series of modifications in HS GAG chains, enormous structure heterogeneity can be generated. Thus, both the modifications of HS GAG chains as well as the nature of the core proteins can potentially play a role in the specificity and function of HSPGs. *In vitro* biochemical studies have demonstrated that HSPGs play critical roles in various cellular processes such as cell adhesion, neurite outgrowth, angiogenesis, tumorigenesis and tissue repair mechanisms (reviewed by Bernfield et al., 1992; David, 1993; Kjellen and Lindahl, 1991; Yanagishita and Hascall,1992). In the context of signal transduction, HSPGs have been implicated to function as co-receptors for a number of growth factors, internalization of receptors and transport of signaling molecules (Salmivirta et al., 1996). However, the importance of these molecules in development remains to be elucidated.

Wingless (Wg) encodes a protein of the Wnt family and acts as a critical regulator in many developmental processes both during embryonic and larval development (reviewed by Siegfried and Perrimon, 1994). Genetic studies in *Drosophila*, in combination with biochemical studies, have led to the identification of a number of downstream molecules required for Wg signaling (for reviews, see Cadigan and Nusse, 1997; Cox and Peifer, 1998; Dale, 1998). According to a current model, Wg protein bind to the seven transmembrane receptor, *Drosophila* Frizzed 2 (Dfz2), and transduces its signaling to several downstream components resulting in the accumulation of cytoplasmic Armadillo (Arm). Arm subsequently translocates into the nucleus to form a complex with LEF/TCF transcription factors. Together Arm and TCF regulate the expression of many downstream target genes such as *engrailed (en)* in the embryonic epidermis, *distalless (dll)* and genes of the *achaete scute complex* (*acs*) in the wing.

Genetic studies have demonstrated that Wg can exert both short-range and long-range effects during the development of the embryo and imaginal discs (Bejsovec and Martinez Arias, 1991; Struhl and Basler, 1993; Hoppler and Bienz, 1995; Lawrence et al., 1996). For example, at the wing margin, *wg* is expressed in a narrow strip of four to five cells wide straddling the dorso-ventral (D/V) boundary, and Wg proteins form a gradient over a distance of up to 25 cell diameters (Zecca et al., 1996, Neumann and Cohen, 1997; Cadigan et al., 1998). Wg acts as a short-range organizer to activate the expression of several genes adjacent to the wing margin, including members of the *acs* complex. Wg also functions as a morphogen to directly activate the transcription of several target genes such as *dll* in a Wg-concentration dependent manner (Zecca et al., 1996, Neumann and Cohen, 1997). However, what factors govern the shape of the Wg morphogen gradient and how the activity of Wg is regulated is not clear. Recent experiments have proposed that the expression of the Wg receptor *Dfz2*, which is negatively regulated by Wg signaling, plays a role in shaping the gradient of Wg activity. Overexpression of *Dfz2* in wing discs was found to increase Wg range, possibly as a result of the stabilization of Wg when it is associated to Dfz2 (Cadigan et al., 1998).

Biochemical studies have shown that both Wg and its mouse homolog Wnt-1 are high-affinity Heparin-binding proteins (Bradley and Brown, 1990; Reichsman et al., 1996). Both Wg and mouse Wnt-1 are poorly secreted and usually tightly associated to the cell surface as well as the ECM (Papkoff and Schryver, 1990; Bradley and Brown, 1990; Gonzalez et al., 1991; van den Heuvel et al., 1993; Reichsman et al., 1996). Further, the

tightly associated Wg proteins on the cell surface can be released by addition of exogenous Heparin (Bradley and Brown, 1990; Reichsman et al., 1996). These characteristics imply that HSPGs might be important factors in Wg function. Indeed, in tissue culture experiments, Wg signaling can be inhibited by removal of HS with Heparinase or by treatment of cells with sodium perchlorate, a competitive inhibitor that blocks the sulfation of proteoglycans (Reichsman et al., 1996). Further, we and others have recently reported that in the absence of *surgarless* (*sgl*) gene activity, which encodes a *Drosophila* homolog of UDP-glucose dehydrogenase, Wg signaling is defective (Binari et al., 1997; Haecker et al., 1997; Haerry et al., 1997). This enzyme is required for the formation of Glucuronic acid (GlcA). Because GlcA is required for the formation of HS, CS and dermatan sulfate (DS), it is not known what class of proteoglycans are involved for Wg signaling.

In this paper, we report the characterization of *sulfateless (sfl)*, which encodes a *Drosophila* homolog of vertebrate HS N-deacetylase/N-sulfotransferase (NDST). In mammalian cells this enzyme is required specifically for the modification of HS GAGs but not CS and DS GAGs. We demonstrate that Sfl is required for Wg signaling in a number of tissues during both embryonic and imaginal discs development. These results provide evidence that HSPGs are involved in Wg signaling and that HSPGs play non-redundant roles with other types of proteoglycans in the context of Wg signaling. Further, we provide evidence that the product of *dally* gene, a *Drosophila* homolog of Glypican (Nakato et al., 1995), encodes the protein component of the HSPG involved in Wg signaling. Interestingly, we find that *dally* is co-expressed with *Dfz2*, and is repressed by Wg signaling as observed for *Dfz2* (Cadigan et al., 1998). Altogether, our results suggest that HSPGs play a key role in modulating Wg activity and that Dally functions as a co-receptor for Wg in regulating both short and long-range activities of Wg.

(b) Results

Sfl is involved in Wg signaling during embryonic development.

sfl was identified in a genetic screen to characterize the maternal effects of zygotic lethal mutations (Perrimon et al., 1996; and Experimental Procedures for details). Homozygous *sfl* mutant animals derived from heterozygous mothers die at the third instar larval or early pupal stages. However, homozygous *sfl* mutant embryos derived from females lacking germline *sfl* activity (referred to as *sfl* null embryos throughout the text) die with a segment polarity phenotype. The *sfl* maternal effect is completely paternally

rescuable indicating that *sfl* is expressed at least both during oogenesis and early embryonic development.

The cuticle phenotype of *sfl* null embryos is reminiscent of the phenotypes exhibited by either mutations in *wg* or *hh*. We further examined the expression of *wg* mRNA, and Wg and En proteins in *sfl* null mutant embryos. The expression patterns of *wg* and *en* in these embryos are reminiscent of those observed in either *wg* or *hh* null mutants (DiNardo et al., 1988; van den Heuvel et al., 1993; Yoffe et al., 1995; Manoukian et al., 1995; Alcedo et al., 1996), and thus are consistent with a role for *sfl* in either Wg and/or Hh signaling. It is worthy to note that in *sfl* null embryos, Wg is nearly completely absent when En begin to fade during late stage 9 in *sfl*. This result suggests that in *sfl* null embryos Wg proteins may have diffused in the extracellular space or have been degraded upon their secretion.

To address more specifically whether Wg signaling is perturbed in the absence of *sfl* activity, we examined the development of several other embryonic tissues that require Wg activity. These include the development of the stomatogastric nervous system (SNS) (González-Gaitán and Jäckle, 1995) and the second midgut construction (Bienz, 1994). Examination of *sfl* null embryos revealed that the SNS invaginations do not occur properly and resemble those found in weak *wg* mutants. Further, in *sfl* null embryos, the second midgut constriction does not form as observed in *wg* mutants. Altogether, our results implicate a requirement for Sfl activity in Wg signaling during embryonic development.

Sfl is required for both short and long-range activities of Wg

To further substantiate a requirement for Sfl activity in Wg signaling, we analyzed the effect of *sfl* mutations during wing imaginal disc development. Wg is required for D/V patterning and acts as a short-range inducer to activate the expression of several genes such as *neuralized* (*neu*) at the wing margin (Phillips and Whittle, 1993; Couso et al., 1994). Wg also functions as a morphogen to directly activate the transcription of several target genes including *dll* in a concentration dependent manner (Zecca et al., 1996, Neumann and Cohen, 1997). *sfl* homozygous mutant animals derived from heterozygous mothers die as late larvae to early pupae. In *sfl* mutant wing discs, the expression of *neu*, detected with the A101 marker, is abolished, suggesting that Sfl is required for short-range Wg activity. Further, the expression of Dll is strikingly reduced in *sfl* mutant discs reflecting a role of Sfl in long-range Wg signaling. Consistent with these results, clones of *sfl* mutant cells

induced at the first instar stage generate severe wing margin defects, a phenotype also seen with *wg* or *disheveled (dsh)* mutant clones (Couso et al., 1994). Interestingly, when induced at a later stage (late second instar), a number of *sfl* mutant clones exhibit partial wing margin defects. This observation indicates that *sfl* does not act in a completely cell autonomous manner at that stage, possibly as the result of perdurance of the wild type Sfl product. Altogether, our results strongly argue that Sfl is required for both short and long-range Wg activities in the wing imaginal disc.

sfl encodes heparan sulfate N-deacetylase/N-sulfotransferase

To molecularly characterize the gene associated with *sfl* mutations, we cloned the genomic DNA flanking the P-element insertion associated with *sfl*$^{l(3)03844}$ following plasmid rescue. A DNA subfragment adjacent to the P-element insertion was used to probe a Northern blot and a 5.0 kb transcript was identified. This DNA subfragment was then used to screen a 0-4 hr embryonic cDNA library. A 4.89 kb full-length cDNA clone was isolated and found to encode a conceptual protein of 1048 amino acid residues in length.

Three lines of evidence suggest that the cDNA we have isolated corresponds to *sfl*. First, the sequence of the cDNA, when compared to the genomic site of insertion of the P-element, revealed that the P-element is inserted 686 bp upstream of a putative ATG start codon in the untranslated 5' region of the cDNA. This suggests that the P-element insertion may disrupt the *sfl* transcript. Second, Northern blots probed using a *sfl* cDNA detect no *sfl* maternal transcripts in 0-1.5 hr embryos derived from females with *sfl* germline clones, indicating that the P-element insertion disrupts the *sfl* transcript. Third, *in vitro* transcribed *sfl* RNAs injected into marked *sfl* mutant embryos were able to rescue the embryonic null mutant phenotype derived from *sfl* germline clones.

A search of the protein sequence databases revealed that the putative protein deduced from *sfl* cDNA has striking homology with vertebrate HS/Heparin N-deacetylase/N-sulfotransferase (NDST) (Hashimoto et al., 1992; Orellana and Hirschberg, 1994; Eriksson et al. 1994; Aikawa and Esko, 1999), which are essential for the modification of HS/Heparin polysaccharide chains. Three isoforms of NDST have been identified in vertebrates and these enzymes all exhibit N-deacetylase/N-sulfotransferase activity with certain differences in expression. As shown in Figure 3D, while the 160 N-terminal amino acid residues are distinct from rat NDST1, the remaining sequences of Sfl is highly homologous with an overall identity of 51%. Similar amino acid identities were

observed between Sfl and vertebrate NDST2 and NDST3 (data not shown), raising the possibility that Sfl may represent a prototype enzyme of NDST (see discussion).

dally, a member of the Glypican family of HSPG, is new segment polarity gene

The analysis of *sfl* implicates that HS GAGs, but not other classes of GAGs, are required for proper Wg signaling. Since HS GAGs are attached to various protein cores to form different HSPGs, we searched for candidate genes that could encode the protein component of the HSPG. The *Drosophila* Glypican homolog *dally* appeared as an excellent candidate because a previous analysis has reported that flies homozygous for hypomorphic *dally* alleles exhibit some wing margin defects (Nakato et al., 1995), a phenotype reminiscent to partial loss of *wg* activity.

As a first step to examine the role of *dally* in Wg signaling, we determined the expression of *dally* mRNAs in embryos by *in situ* hybridization. We found that *dally* is expressed both maternally and zygotically. At early stages *dally* transcripts are uniformally expressed, however, at stage 8 they are enriched in a segmentally repeated pattern. In stage 8 embryos, *dally* transcripts are expressed in three to four cells anterior to *wg*-expressing cells. Interestingly, the stripes of *dally* expression overlap with those of the Wg receptor *Dfz2* (data not shown) (Bhanot et al., 1996) suggesting a role for *dally* in Wg signaling.

To further demonstrate that *dally* is required for Wg signal transduction, we examined the cuticle phenotype of *dally* mutant embryos. All the available *dally* alleles are homozygous viable to some extent, with $dally^{p2}$, $dally^{D188}$ and $dally^{PD1}$ representing the strongest alleles available (Natako et al., 1995; and Experimental Procedures). Since *dally* is maternally expressed, to generate stronger resulting *dally* mutant embryos, we recombined all the stronger *dally* alleles with the FRT^{2A} chromosome to generate females with homozygous *dally* germline clones. *dally* mutant embryos derived from females homozygous for *dally* alleles lacking germline *dally* activity exhibits poorly penetrant cuticle segment polarity defects resembling a partial defect in Wg signaling. However, this phenotype can be significantly enhanced by removal of one copy of *sfl* in the mother, or one copy of *wg* in the embryo. These genetic dosage interactions are consistent with a role of *dally* in Wg signaling.

Because all available *dally* mutants are weak alleles, we used double-stranded RNA (dsRNA) interference as a method to interfere with the activity of the endogenous *dally*

gene (Kennerdell and Carthew, 1998). We injected dsRNAs corresponding to the entire coding region of *dally* into wild type embryos. Embryos injected with the *dally* dsRNAs exhibit a severe segment polarity cuticle defects, similar to those injected with *wg* or *fz* + *Dfz2* dsRNAs (Kennerdell and Carthew, 1998). No significant defects were observed in a control experiment injected with buffer. This result, together with the genetic interaction observed between *sfl* and *wg* strongly argues that *dally* is a new segmentation polarity gene and that it is required at least for Wg signaling in the embryo.

Dally is required for both short and long-range signaling of Wg in wing imaginal discs.

To further examine the role of *dally* in Wg signaling, we analyzed the function of *dally* during wing imaginal disc development. Consistent with previous reports (Nakato et al., 1995), we observed that only 3% of homozygous *dally* animals, that carry the weak alleles $dally^{P2}$ or $dally^{D188}$, can be recovered and exhibit wing margin defects. This frequency can be increased by 2 to 3 folds when a single copy of *wg* is removed. The enhancement of the wing defects of *dally* mutant by a reduction in Wg activity suggests that Dally plays a role in Wg signaling at the wing margin.

To determine whether Dally cooperates with the Wg receptor Dfz2 in wing patterning, we tested whether *dally* mutations can enhance a loss of function *Dfz2* phenotype. Ectopic expression of a dominant negative form of *Dfz2* (Dfz2N), that encodes the first extra-cellular domain and the first transmembrane domain, has been shown to block Wg signaling, probably by binding to Wg in a non productive manner. When Dfz2N is expressed ectopically using the Gal4 line C96, which drives expression in the presumptive wing margin, flies develop partial margin defects. However, this phenotype is dramatically enhanced in homozygous *dally* mutants suggesting that *dally* potentiates Wg signaling. Importantly, ectopic expression of a gain of function Arm protein (Arm s10) (Pan et al., 1997), in which the 54 N-terminal amino acid residues of Arm are deleted, can fully rescue the wing defects, suggesting that Dally acts upstream of Arm. Similar results were obtained when another dominant negative form of Dfz2, GPI-Dfz2N (Cadigan et al., 1998), was used (data not shown). Altogether, these genetic interactions are consistent with a role of Dally in Wg signaling and suggest that Dally may act together with Dfz2 in Wg reception.

If *dally* acts together with Dfz2 to mediate Wg signaling, Dally may also be required for the other functions of Dfz2 in Wg signaling. In the wing blade, Dfz2 is involved in shaping the gradient of Wg distribution and determine the response of cells to Wg. Uniform *Dfz2* overexpression in the wing pouch leads to ectopic bristles formation in the wing blade, most likely reflecting the activation of Wg signaling above its normal level (Cadigan et al., 1998). As observed previously, ectopic expression of *Dfz2* driven by the Gal4 line 69B resulted in wings with ectopic bristles. In a *dally* mutant background, the formation of ectopic bristles was drastically reduced suggesting that a mutation in *dally* blocks the activity of Dfz2. Further, we examined the effect of *dally* on the expression of *dll*, which is normally expressed at high levels close to *wg*-expressing cells and form a gradient at further distance. Ectopic expression of *Dfz2* expanded the expression levels of Dll much further from the D/V border than in the control. The expression of Dll in *dally* mutant where *Dfz2* is overexpressed is significantly reduced, suggesting that a reduction of *dally* activity reduces Wg signaling.

It has been previously demonstrated that misexpression of *Dfz2* alters Wg distribution by increasing its stability (Cadigan et al., 1998). We tested whether a reduction in *dally* gene activity would affect the formation of Wg/Dfz2 signaling complexes. If this is the case, one might expect that Wg distribution, in wing imaginal discs overexpressing *Dfz2* and mutant for *dally*, would be affected. Wg distribution was compared in *UAS-Dfz2/69B-Gal4* and *UAS-Dfz2 dallyP2/69B-Gal4 dallyP2* wing discs. A significant reduction in Wg protein distribution was observed in a *dally* mutant background, suggesting that a reduction in *dally* activity destabilizes Wg/Dfz2 complexes. An alternative model to explain the effect of *dally* on Wg distribution, is that Dally is involved in the transport of Wg proteins.

Dally expression is similar to Dfz2 and inhibited by Wg signaling

Previous studies have demonstrated that Wg signaling inhibits *Dfz2* expression in the wing imaginal disc (Cadigan et al., 1998). Similarly, when we examined the expression of *dally* in wing discs, we found that *dally* mRNAs are expressed in a pattern very similar to *Dfz2* mRNAs, with lowest level at the D/V boundary. To determine whether the expression pattern of *dally* is negatively regulated by Wg signaling, as previously shown for *Dfz2* (Cadigan et al., 1998), we expressed a dominant negative form of dTCF ($dTCF^{DN}$) to block Wg signaling. When Wg signaling activity is repressed by ectopic expression of

dTCF^{DN} in the Patched (Ptc) domain, *dally* transcription is strikingly increased. Thus, as in the case of *Dfz2*, *dally* expression is down-regulated by Wg signaling.

(c) Discussion

Sfl and Dally are involved in Wg signaling.

We have characterized a novel segmentation polarity gene, *sfl*, that encodes an enzyme specifically required for the modification of HS GAGs. In the absence of both maternal and zygotic *sfl* gene activity, embryos develop with cuticle patterning defects similar to those observed in either *wg* and/or *hh* mutant embryos. *sfl* embryos also show defects reminiscent to loss of *wg* activity in other tissues such as the SNS and midgut. In addition, we found that *sfl* is required for both short and long-range Wg signaling in the wing imaginal disc. All the defects observed in *sfl* null embryos are identical to those associated with mutations in the *sgl* gene, which encodes another enzyme involved in the biosynthesis of various proteoglycans (Hempel et.al, 1994). However, unlike Sgl which is involved in the formation of HS, CS and DS GAGs, Sfl activity is predicted to be specific to the biosynthesis of HS. Thus, our results provide the first genetic evidence that HSPGs, rather than other proteoglycans, are required for Wg signaling during both embryonic and larval development.

We have identified the product of the *dally* gene, a member of the *glypican* family, as the protein core of the HSPGs involved in Wg signaling. Several observations support the model that Dally function in Wg reception. First, disruption of *dally* function in the embryo, either by mutations or by dsRNA interference, is associated with patterning defects reminiscent to loss of *wg* activity. Second, a reduction in *dally* gene activity in the wing exhibits margin defects, a phenotype which can be further enhanced by either a reduction in Wg activity, or following ectopic expression of a dominant negative form of the Wg receptor Dfz2. Third, *dally* mutations block the activity of ectopic expression of *Dfz2* and affect Wg distribution in the wing disc. Finally, *dally* is expressed in a pattern that resembles the expression of *Dfz2* both in the embryo and wing disc, and, as observed for *Dfz2*, is down-regulated by Wg signaling.

HS GAG modifications and Wg signaling.

In vertebrates, HS is a ubiquitous GAG that is closely related to Heparin, a GAG expressed *in vivo* solely as a proteoglycan within the granules of mast cells and basophils

(reviewed by Bernfield et al., 1992; Kjellén and Lindahl, 1991). Biosynthesis of HS and Heparin are similar in vertebrates and invertebrates, and are initiated by formation of a polysaccharide chain consisting of multimers of D-glucuronic acid b1,4 - N-acetyl-D-glucosamine a1,4 (GlcA-GlcNac) disaccharides. Subsequently, several modifications including N-deacetylation and N-sulfation, Uronosyl C5-epimerization, 2-0-sulfation, 6-0-sulfation and 3-0-sulfation occur (reviewed by Kjellén and Lindahl, 1991, Lindahl et al., 1998). HS N-deacetylase/N-sulfotransferase (NDST) catalyzes N-deacetylation and N-sulfation, which is a coupled reaction and is the first and key step to initiate further modification reactions of HS/Heparin. Sfl protein is strikingly homologous with all three isoforms of vertebrate NDST. These proteins are homologous enzymes with certain differences in expression and possibly in specificity (Hashimoto et al., 1992; Eriksson et al., 1994; Orellana and Hirschberg, 1994; Aikawa and Esko, 1999). These enzymes have been suggested to be derived from an ancestral gene (Eriksson et al., 1994). Since *sfl* is expressed ubiquitously throughout embryonic development, it is likely that Sfl represents a homolog of vertebrate NDST or a prototype enzyme of vertebrate NDST. The nature of Sfl implies that in *sfl* null mutants, modification of HS GAGs do not occur. Since sulfation and epimerization of HS/Heparin molecules provide structural identity and negative charges which are critical for its interaction with proteins, loss of Sfl activity will result in the generation of unmodified HS/Heparin, and thus impair the normal function of HSPGs. Based on these considerations, we propose that in the absence of Sfl, Wg signaling is perturbed due to the abnormal biosynthesis of HSPGs.

If the general synthesis of HSPGs is perturbed in *sfl*, then it is expected that other signaling pathways, which utilize HSPGs, may also be disrupted in the absence of Sfl. Consistent with this hypothesis, we have observed that both FGF signaling pathways in *sfl* are defective as seen in the embryos with mutations of *heartless* and *breathless*, which encode homologues of FGF receptors (Lin et al., submitted). In addition, we have also observed that Hh signaling is affected in *sfl* mutant (unpublished observations). The defects of other signaling pathways in *sfl* mutant argue that HSPGs are a major class of proteoglycans required for signaling of many key growth factors in development.

*Dally/Glypican HSPG and Wg/*Wnt *signaling*

Our genetic data implicate Dally as a molecule involved in Wg signaling. However, because all the *dally* mutations isolated to date are weak alleles (see Experimental Procedures), we cannot completely rule out the possibility that at least another HSPG may

play a partially redundant function with Dally in Wg signaling. Thus, until the complete loss of function *dally* phenotype has been characterized and compared to the *wg* null phenotype, the possibility remains that another HSPG is also involved in Wg signaling. Interestingly, a search of data base revealed that there is another *Drosophila* member of the Glypican family in the EST database. Another possibility is that Syndecan, which is with Glypican the other major cell surface HSPGs, contribute to some extent to Wg reception. However, we think that this is unlikely because Syndecans and Glypicans differ in many ways. Syndecans bear both CS and HS GAGs while Glypicans are attached exclusively to HS GAGs. Further, Glypicans share several unique features that include a characteristic pattern of 14 highly conserved cysteine residues, 2 to 3 Ser-Gly HS GAG attachment sequences near the C-terminus, and a C-terminal sequence involved in the formation of a GPI-linkage to the membrane (Veugelers and David, 1998), suggesting that they have different functions from Syndecans. Finally, we find that in the embryo *Dsyndecan* is highly expressed in the mesoderm and trachea suggesting that it may be involved in FGF signaling (unpublished results).

Because most signaling mechanisms are evolutionary conserved, we expect that Glypicans may also play a role in mammalian Wnt signaling pathways. Consistent with this model, most vertebrate glypicans are expressed in tissues, for example brain and kidney, where Wnt family members are expressed. It has been demonstrated that HSPG is required for the maintenance of mouse Wnt-*11* expression in the ureter tips of the developing kidney. Further, treatment of cultured kidney rudiments with either chlorate or Heparinase III resulted in the reduction of Wnt *11* expression, suggesting that HSPG is required for autocrine signaling activity of Wnt 11. Interestingly, mouse K-glypican is expressed in a very similar pattern to Wnt 11 in a developing kidney (Watanabe et al., 1995). Further, mutation in human Glypican-3 are responsible for the Simpson-Golabi-Behmel Syndrome, a disease associated with prenatal and postnatal overgrowth and a high incidence of neuroblastomas and Wilm's tumors (Pilia et al., 1996). Since many Wnts are required for cell differentiation and growth control in brains and kidney, it is possible that this disease is associated with a defect in Wnt signaling.

Possible functional diversity of Dally in different tissues.

It has been reported that Dally regulates the activity of decapentaplegic (dpp), a member of the TGF-b superfamily (Jacson et al., 1997). A reduction in dpp levels enhance

the defects associated with dally mutations in the eye, antenna and genitalia. Further, additional copies of dpp rescues the defects in these tissues. However, in the wing, a reduction in dpp activity has been reported to rescue the incomplete wing vein V and notching defects associated with dally mutants. In addition, extra copies of dpp can enhance the dally wing defects. These genetic interactions have led to the hypothesis that Dally regulates Dpp activity, however the nature of this interaction is unclear and could be indirect. Dpp activity in the wing is controlled by Hh signaling pathway which itself requires HSPGs (The et al., submitted), raising the possibility that the interaction between Dally and Dpp reflects a role for Dally in Hh signaling. The alternative model is that Dally plays a direct role in Dpp reception. In this scenario, the function of Dally would be tissue specific since we have no evidence for a function of HSPGs in the early function of Dpp in the establishment of D/V embryonic polarity. Tissue specific effects of Dally could be generated either through tissue specific expression of dally during development or tissue specific modification of the HS GAG chains linked to the Dally protein core. Although the dynamic expression pattern of dally that we have detected is more reminiscent of a function of dally in Wg signaling rather than Dpp, our analysis does not rule out that this HSPG could play a role either Dpp or Hh signaling, especially since we have not examined Dally protein distribution. Finally, there is a number of biochemical and genetic evidence supporting that the model that specific HS GAG decorates the cell surface. In vertebrate, a number of sulfotransferase have been shown to be differentially expressed in various tissues (Kimato's papers; Beddington). Finally, the Drosophila gene pipe, which is involved in D/V patterning in the embryo, encodes a HS 2-O sulfotransferase (Sen et al., 1998) that is expressed in ventral follicle cells. Future studies will be required to clarify the function of Dally in Dpp signaling.

The role of Dally in Wg signaling

Previous genetic and biochemical evidence have revealed that binding of Wg to the extracellular domain of the seven transmembrane protein Dfz2 triggers signaling as revealed by a modulation in the level of cytoplasmic Arm. In this example, we present genetic evidence that the Glypican protein Dally is also involved in the activation of Dfz2. Because Dally is an extracellular GPI-linked protein, GPI-linked protein has been demonstrated to be located in lipid rafts, which are microdomains in the plasma membrane rich in sphingolipids, cholesterol and GPI-anchored proteins [Simons, 1997 #23], we envision two possible models for its role in Wg/DFz2 signaling. The first one is a "co-

receptor" model, whereby Dally works together with Dfz2 to trap and stabilize Wg. The second one is a "transport" model, whereby Dally is involved in the movement of Wg through field of cells.

The co-receptor model for the role of Dally in Wg signaling is similar to the proposed role of HSPGs in other growth factor signaling pathways such as FGF. In this case, the binding of growth factors to abundant but low affinity HS GAGs on the cell surface limits the free diffusion of the ligand from three to two dimensions, thereby increasing its local concentration and the probability that it will interact with less abundant, high affinity signaling receptor (Schlessinger et al., 1995). According to this model, we propose that the HS GAG chains of Dally trap secreted Wg. The nature of the GPI anchor may facilitate lateral movement of Wg/Dally complexes such that Wg molecules can encounter less abundant Dfz2 receptor molecules. Interaction of Dally/Wg with Dfz2 may further stabilize it or form an active ligand/receptor complex. Interestingly, *dally* and *Dfz2* are expressed and regulated in a very similar manner. In the case of *Dfz2* it has been suggested that the gradient of *Dfz2* expression, which is at its highest in cells that receive little Wg, is important for shaping the Wg morphogen gradient. Our data with Dfz2 are consistent with Dally also playing a role in the stabilization of Wg protein at the cell surface because when Dally levels are reduced, Wg distribution stabilized by Dfz2 is also reduced. This model is also consistent with our observation that in either *sgl* and *sfl* mutant embryos, overexpression of Wg can compensate for the loss of HS GAG and rescue the embryonic cuticle segmentation defects (Haecker et al., 1997; data not shown). It is also evident that Wg protein is absent in the either sgl Haecker et al.,) and *sfl* mutant embryos at the time En protein are still quite normal. Thus, according to the co-receptor model, a combination of Dally and Dfz2 is needed to achieve optimal activation of the transducing Dfz2 receptor. Biochemical experiments that address whether Dfz2 and Dally physically associate should help test this model.

The second model propose a role for Dally in Wg movement through field of cells rather than playing a direct role in activation of Dfz2. The mechanism underlying the movement of Wg/Wnt molecules through tissues is not understood, and it is not clear whether the gradients of activities that these proteins trigger are established through diffusion of the secreted factors in the extracellular space or through transport mechanisms that involve for example vesicle-like structures which are endocytosed and/or transcytosed. There is evidence that Wnts can be secreted in the extracellular space as well as transported

through cells. Wnt proteins are poorly secreted in the extracellular space and bind the extracellular matrix tightly [Bradley, 1990 #39][Gonzalez, 1991 #29][Reichsman, 1996 #38]. These features suggest that these proteins may not be able to freely diffuse and thus raise the questions of how many cells away they are able to act. Immunohistochemical analyses have detected Wg protein up to 25 cell diameters from the wing margin, which is the site of Wg synthesis in this tissue [Cadigan, 1998 #31]. Thus, Wg proteins can be found far away from *wg*-expressing cells raising the issue of the mechanism underlying the movement of Wg through cells. Interestingly, electron microscopy studies detect little Wg protein free in the extracellular space and Wg is present in vesicle like structures [Gonzalez, 1991 #29]. The presence of Wg in vesicles is endocytosis dependent and is not detected in shibire mutants which are defective in endocytosis [Bejsovec, 1995 #26]. This observation has led to the model that Wg proteins are transported through cells in vesicle-like structures, a process referred to as transcytosis [Gonzalez, 1991 #29]. Thus, one possibility is that Dally plays a direct role in the movement/transcytosis of Wg through cells, pohapes through the GPI-ancored with is located lipid rafts [Simons, 1997 #23]. According to this model, Wg proteins are transported from cells to cells following its association with the HSPG Dally. Our results however, do not allow us at this time to implicate Dally in Wg movement through field of cells.

(e) Experimental Procedures

Fly stocks

sfl alleles: A single *P[lacZ, ry$^+$]* element insertion, *l(3)03844* (Spradling et al., 1995), located at 65C1-2 was identified in a screen for maternal effects of zygotic lethal mutations (Perrimon et al. 1996). We found that *sfl*$^{l(3)03844}$ fails to complement *Df(3L)ZN47* that deletes the 64C to 65C region of the third chromosome. This result is consistent with the location of a zygotic lethal mutation in the 65C1-2 region. Further, we mobilized the P-element insertion associated with *sfl*$^{l(3)03844}$ using the *y w; D2-3, Sb/TM6* strain (Robertson et al. 1988). Out of 60 excision lines analyzed, 9 were viable in trans with *Df(3L)ZN47*, indicating that the P-element insertion at 65C is associated with zygotic lethality. These revertant lines fully rescued the maternal effect segment polarity phenotypes associated with *l(3)03844*. One additional *sfl* allele (*sfl*^{9B4}) was identified in a large EMS screen for

maternal effects of zygotic lethal mutations (N.P., L. Cunningham and C. Arnold, unpublished data). Both *sfl*$^{l(3)03844}$ and *sfl*^{9B4} show similar maternal effect phenotypes.

dally alleles: dally^{P1}, *dally*^{P2}, *dally*^{D188} are hypomorphic alleles and were a gift from S. Selleck (Nakato et al. 1995). dally^{PD1} is a new hypomorphic allele generated by excision of the P-element associated with *dally*^{P1} **(REF? who made this**?). In an attempt to generate stronger loss of function dally allele, we generated following P-element excisions, a number of new *dally* alleles. However, none were stronger than the original ones described by Nakato et al. (1995).

Other stocks: UAS-Dfz2 and UAS-GPI-Dfz2 lines were obtained from R. Nusse (Cadigan et al., 1998). A strong third chromosome UAS-Dfz2 line was used for all experiments. UAS-Dfz2N was obtained from R. Carthew (Zhang and Carthew, 1998). A strong second chromosome UAS-Dfz2N line N33 was used for all the experiment. The UAS-arm^{act} line S10 and the UAS-dTCF^{DN} line DN4 expressing a deletion from of Arm (Pai et al.1997) or dTCF (van de Wetering et al., 1997) were obtained from M.Peifer. The Gal4 lines used were 69B-Gal4 (Brand and Perrimon, 1993), Ptc-Gal4 (Johnson et al.,1995). C96 Gal4 was identified from a collection of PGawB insertions generated by K. Kaiser. *wg*^{IG22} is a *wg* null allele (Bejsovec and Wieschaus, 1993). A101 (*neulaczZ*) is described in Usui and Kimura (1992).

Maternal effect phenotype of sfl and dally

Females with germline clones were generated using the autosomal "FLP-DFS" technique (Chou and Perrimon, 1996). *sfl FRT*^{2A}*/TM3, Sb* females were mated with males of the genotype *y w FLP*^{22}*/+, FRT*^{2A} *P[ovo*^{D1}*]/TM3, Sb*. The resulting progeny were heat shocked at 37°C for 2 hrs during the larval stages, and *y w FLP*^{22}*/+; sfl FRT*^{2A}*/FRT*^{2A} *P[ovo*^{D1}*]* females carrying *sfl* homozygous germline clones were selected. Females with germline clones of *dally* were generated using a *dally FRT*^{2A} recombinant chromosome using the same procedure as for *sfl*.

Generation of somatic wing clones.

For the generation of adult somatic clones in the wing, *y w* hsFLP122; *sfl*$^{l(3)03844}$ *FRT*^{2A}*/TM3, Sb* were crossed with males of *y w, P[y+] FRT*^{2A}*/TM3, Sb*. Larvae from this

cross were heat-shock for 2 hours at first or second instar. Adult wings were mounted in Euparal for observation.

Antibody staining and in situ hybridization

Fixation of embryos and antibody staining procedures were performed as described (Patel 1994) . Anti-Wg serum was a gift of S. Cumberledge and used at 1:500 dilution. Anti-En MAb4D9 was used at 1:300 dilution and obtained from Developmental Studies Hybridoma Bank (Patel et al. 1989). Anti-Dll serum was obtained from I. Duncan (Duncan et al., 1998) and used at 1:500. Antibody against the Crumbs protein was used at 1:50 and obtained from E. Knust (Tepass and Knust 1993).

In situ hybridizations of whole-mount embryos and imaginal discs were done with RNA probe (Lehmann and Tautz 1994). *wg* digoxigenin-labeled DNA probe was prepared from a *wg* cDNA subcloned in the p^{sp65} plasmid. dally cDNA was obtained from S. Selleck (Natako et, al 1995). Dfz2 cDNA was obtained from R. Nusse (Bhanot et al., 1996).

Molecular biology

Genomic DNAs flanking the *sfl*$^{l(3)03844}$ P-element insertion was obtained by plasmid rescue in *E. coli* (Cooley et al. 1988) . To isolate *sfl* cDNAs, we screened a 0-4 hr embryonic cDNA library (Brown and Kafatos 1988) with random primed probes generated from genomic DNA fragments from plasmid rescue. One full-length 4.89 kb *sfl* cDNA was isolated. DNAs were sequenced by Taq-polymerase cycle sequencing and an automatic sequencer. To define the P-element genomic insertion site, rescued plasmids from *sfl* flies were also sequenced using a primer derived from the P-element.

Northern blots of total RNA were carried out by standard procedures (Sambrook et al. 1989). Probes used are as following: 2.5 kb Xho-EcoRI fragment of *sfl* cDNA and *sgl* cDNA was used as a control (Haecker et al., 1997). Sequence alignment was produced using "DNA star" software.

RNA injection rescue

RNA injection rescue was done as described (Heacker et al., 1997). The *sfl* mutation was recombined with a *trachealess* (*trh*) mutation (Wilk et al., 1996; Isaac and

Andrew 1996) located at 61 C1-2. *trh* mutants are missing tracheas and exhibit defective posterior spiracles which are easily scorable by looking at embryonic cuticles. Females that carry *trh sfl* homozygous germline clones were generated and mated with *sfl/trh* transheterozygous males, respectively. Half of the embryos from this cross are *sfl* mutants and exhibit strong segment polarity phenotypes, while the remaining half of the embryos which are completely paternally rescued are easily identifiable by their defective posterior spiracles. Of 700 injected embryos derived from females with *trh sfl* germline clones, 120 *sfl* mutant embryos (*trh sfl/sfl*) developed scorable cuticle structures, and 15% of them showed evidence of rescue.

dsRNA-mediated genetic interference

The dsRNA synthesis and injection were as described (Kennerdell and Carthew, 1998). The *dally* full length cDNA cloned in pBluescript was used as a template used for RNA synthesis. Both sense and antisense RNA of *dally* were synthesized using T7 and T3 polymerase by MEGAscript kit from Ambion company. A mixture of equal amount of sense and antisense RNA were heated in boiling water for 1 minute and then allowed to cool to room temperature for overnight. A concentration of 1µg/µl of *dally* dsRNA was injected in the posterior domain extending from 50% to 75% egg length at syncitial blastoderm stage. In a control experiment, the injection buffer was used.

We Claim:

1. An assay for identifying an inhibitor of a proteoglycan-dependent signal transduction activity of an extracellular protein, comprising

   (i) providing a reaction mixture comprising

   (a) a glycosyltransferase which is essential to synthesis of a proteoglycan and which selectively regulates the signal transduction activity of the extracellular protein,

   (b) a substrate for the glycosyltransferase, and

   (c) a test agent,

   under conditions wherein the glycosyltransferase converts the substrate to a detectable product in the absence of the test agent; and

   (ii) detecting the conversion of the substrate to product,

   wherein a decrease in the rate of conversion of the substrate to product in the presence of the test compound, relative to its absence, indicates that the test compound is an inhibitor of the glycosyltransferase

2. The method of claim 1, wherein the extracellular protein is a growth factor or cytokine.

3. The method of claim 2, wherein the extracellular protein is hedgehog protein, e.g., a mammalian *hedgehog* protein, e.g., *Desert hedgehog* (*Dhh*), *Sonic hedgehog* (*Shh*) or *Indian hedgehog* (*Ihh*).

4. The method of claim 2, wherein the extracellular protein is a Wnt protein, e.g., a mammalian Wnt protein, e.g., Wnt -1, 2, 3a, 3b, , 4, 5a, 5b, 6, 7a, 7b, 8a, 8b, 10a, 10b, 11, or 12.

5. The method of claim 1, wherein the glycosyltransferase is selected from the group consisting of an N-acetylgalactosaminyltransferase (GalNAc transferase) such as EC 2.4.1.40, EC 2.4.1.92, EC 2.4.1.41; an N-acetylglucosaminyltransferase (GlcNAc transferase) such as EC 2.4.1.101, EC 2.4.1.143, EC 2.4.1.144, EC 2.4.1.155; an N-acetylglucosaminyl-1-phosphate transferase (GlcNAc-1-P transferase) such as EC 2.7.8.15; a Fucosyltransferase (Fuc transferase) such as EC 2.4.1.69, EC 2.4.1.65; a galactosyltransferase (Gal transferase) such as EC 2.4.1.151, EC 2.4.1.90, EC 2.4.1.22, EC 2.4.1.37, EC 2.4.1.45; a Glucosyltransferase (Glc transferase) such as EC 2.4.1.117;

a glucuronyltransferase such as EC 2.4.1.17; a Mannosyltransferase (Man transferase) such as EC 2.4.1.83, EC 2.4.1.109; an oligosaccharyltransferase such as EC 2.4.1.119; and a sialyltransferase (NeuAc transferase) such as EC 2.4.99.4, EC 2.4.99.6, EC 2.4.99.1, EC 2.4.99.3, EC 2.4.99.8.

6. The method of claim 1, wherein the glycosyltransferase is a glycosaminoglycan sulfotransferases and/or glycosaminoglycan deacetylases.

7. The method of claim 1, wherein the glycosyltransferase is a glycosaminoglycan sulfotransferases selected from the group consisting of an aryl sulfotransferase (EC 2.8.2.1), an arylamine sulfotransferase (EC 2.8.2.3), a chondroitin 4-sulfotransferase (EC 2.8.2.5), a UDP-N-acetylgalactosamine-4-sulfate sulfotransferase (EC 2.8.2.7), a desulfoheparin sulfotransferase (EC 2.8.2.8), a galactosylceramide sulfotransferase (EC 2.8.2.11), a heparitin sulfotransferase (EC 2.8.2.12), a chondroitin 6-sulfotransferase (EC 2.8.2.17), a triglucosylalkylacylglycerol sulfotransferase (EC 2.8.2.19), a keratan sulfotransferase (EC 2.8.2.21), an arylsulfate sulfotransferase (EC 2.8.2.22), a heparin-glucosamine 3-O-sulfotransferase (EC 2.8.2.23), and a desulfoglucosinolate sulfotransferase (EC 2.8.2.24).

8. The method of claim 1, wherein the glycosyltransferase is an N-deacetylase (EC 3.1.1.-) / N-sulfotransferase (EC 2.8.2.-).

9. The method of claim 1, wherein the glycosyltransferase is a sulfotransferase which interacts with an Ext protein and/or is essential for hedgehog-specific GAG chains.

10. The method of claim 1, wherein the glycosyltransferase is a sulfotransferase which is essential for Wnt-specific GAG chains, e.g., is essential for synthesis of glypicans which selectively interact with Wnt proteins.

11. The method of claim 1, including the further step of formulating one or more inhibitors identified in the assay in a pharmaceutically acceptable excipient.

12. The assay of claim 1, wherein the reaction mixture is a cell-free protein preparation.

13. The assay claim 1, wherein the reaction mixture comprises a recombinant cell including a heterologous nucleic acid recombinantly expressing the glycosyltransferase.

14. A method of inhibiting the activity of a growth factor or cytokine comprising contacting a cell, *in vitro* or *in vivo*, with an inhibitor identified in the assay of claims 1.

15. A method of inhibiting the activity of a growth factor or cytokine comprising contacting a cell, *in vitro* or *in vivo*, with a compound which specifically inhibits sythesis of GAG chains selective for the factor.

16. The method of claim 14 or 15, wherein the inhibitor is administered to an animal, such as a human or other mammal, to treat a disorder involving unwanted growth factor or cytokine activity.

17. The method of claim 14-16, wherein the inhibitor is a small organic molecule.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(ii) TITLE OF INVENTION:

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Foley, Hoag & Eliot
(B) STREET: One Post Office Square
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02109

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Asc II (text)

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 30-MAR-1999
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Vincent, Matthew P.
(B) REGISTRATION NUMBER: 36,709
(C) REFERENCE/DOCKET NUMBER: HMV-042.01

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 832-1000
(B) TELEFAX: (617) 832-7000

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 652..2889

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTCCAGGCC CCGCCGCGCG TCCCGGGGGC CGGCCCCGCG AGCGCAGGAG TAAACACCGC 60

```
CGGAGTCTTG GAGCCGCTGC AGAAGGGAAT AAAGAGAGAT GCAGGGATTT GTGAGGTTAC    120

GGCGCCCCAG CTGCAAGATG CACTAGCCGG CTGAACCCGG GATCGGCTGA CTTGTTGGAA    180

CCGGAGTGCT CTGCACGGAG AGTGGTGGAT GAGTTGAAGT TGCCTTCCCG GGGCTCATTT    240

TCCACGCTGC CGAGAGGAAT CCGAGAGGCA AGGCAATCAC TTCGTCTTGC CATTGATTGG    300

GTATCGGGAG CTTTTTTTTT CTCCCCTCTC TCTTTCTTTT CCTCCGTCTT GTTGCATGCA    360

AGAAAATTAC AGTCCGCTGC TCGCCCGCCC TGGGTGCGAG ATATTCAGCC CCGCTCTCTC    420

CCGTGCATTG TGCAACCCAA AGATGAAAGA CCGAAGGGGA GAAAGTTAAA GAAATCGCCC    480

ACATGCGCTG GATCAGTCCA CGGCTTGGGG AAAGGCATCC AGAGAAGGTG GGAGCGGAGA    540

GTTTGAAGTC TTTACAGGCG GGAAGATGGC GGACTGGAGC TGAAAGTGTT GATTGGGAAA    600

CTTGGGTGAT TCTTGTGTTT ATTTACAATC CTCTTGACCC AGGCAGGACA C ATG CAG    657
                                                         Met Gln
                                                           1

GCC AAA AAA CGC TAT TTC ATC CTG CTC TCA GCT GGC TCT TGT CTC GCC    705
Ala Lys Lys Arg Tyr Phe Ile Leu Leu Ser Ala Gly Ser Cys Leu Ala
        5                   10                  15

CTT TTG TTT TAT TTC GGA GGC TTG CAG TTT AGG GCA TCG AGG AGC CAC    753
Leu Leu Phe Tyr Phe Gly Gly Leu Gln Phe Arg Ala Ser Arg Ser His
    20                  25                  30

AGC CGG AGA GAA GAA CAC AGC GGT AGG AAT GGC TTG CAC CAC CCC AGT    801
Ser Arg Arg Glu Glu His Ser Gly Arg Asn Gly Leu His His Pro Ser
35                  40                  45                  50

CCG GAT CAT TTC TGG CCC CGC TTC CCG GAG CCT CTG CGC CCC TTC GTT    849
Pro Asp His Phe Trp Pro Arg Phe Pro Glu Pro Leu Arg Pro Phe Val
                55                  60                  65

CCT TGG GAT CAA TTG GAA AAC GAG GAT TCC AGC GTG CAC ATT TCC CCC    897
Pro Trp Asp Gln Leu Glu Asn Glu Asp Ser Ser Val His Ile Ser Pro
            70                  75                  80

CGG CAG AAG CGA GAT GCC AAC TCC AGC ATC TAC AAA GGC AAG AAG TGC    945
Arg Gln Lys Arg Asp Ala Asn Ser Ser Ile Tyr Lys Gly Lys Lys Cys
        85                  90                  95

CGC ATG GAG TCC TGC TTC GAT TTC ACC CTT TGC AAG AAA AAC GGC TTC    993
Arg Met Glu Ser Cys Phe Asp Phe Thr Leu Cys Lys Lys Asn Gly Phe
    100                 105                 110

AAA GTC TAC GTA TAC CCA CAG CAA AAA GGG GAG AAA ATC GCC GAA AGT   1041
Lys Val Tyr Val Tyr Pro Gln Gln Lys Gly Glu Lys Ile Ala Glu Ser
115                 120                 125                 130

TAC CAA AAC ATT CTA GCG GCC ATC GAG GGC TCC AGG TTC TAC ACC TCG   1089
Tyr Gln Asn Ile Leu Ala Ala Ile Glu Gly Ser Arg Phe Tyr Thr Ser
            135                 140                 145

GAC CCC AGC CAG GCG TGC CTC TTT GTC CTG AGT CTG GAT ACT TTA GAC   1137
Asp Pro Ser Gln Ala Cys Leu Phe Val Leu Ser Leu Asp Thr Leu Asp
```

150 155 160

AGA GAC CAG TTG TCA CCT CAG TAT GTG CAC AAT TTG AGA TCC AAA GTG 1185
Arg Asp Gln Leu Ser Pro Gln Tyr Val His Asn Leu Arg Ser Lys Val
165 170 175

CAG AGT CTC CAC TTG TGG AAC AAT GGT AGG AAT CAT TTA ATT TTT AAT 1233
Gln Ser Leu His Leu Trp Asn Asn Gly Arg Asn His Leu Ile Phe Asn
180 185 190

TTA TAT TCC GGC ACT TGG CCT GAC TAC ACC GAG GAC GTG GGG TTT GAC 1281
Leu Tyr Ser Gly Thr Trp Pro Asp Tyr Thr Glu Asp Val Gly Phe Asp
195 200 205 210

ATC GGC CAG GCG ATG CTG GCC AAA GCC AGC ATC AGT ACT GAA AAC TTC 1329
Ile Gly Gln Ala Met Leu Ala Lys Ala Ser Ile Ser Thr Glu Asn Phe
215 220 225

CGA CCC AAC TTT GAT GTT TCT ATT CCC CTC TTT TCT AAG GAT CAT CCC 1377
Arg Pro Asn Phe Asp Val Ser Ile Pro Leu Phe Ser Lys Asp His Pro
230 235 240

AGG ACA GGA GGG GAG AGG GGG TTT TTG AAG TTC AAC ACC ATC CCT CCT 1425
Arg Thr Gly Gly Glu Arg Gly Phe Leu Lys Phe Asn Thr Ile Pro Pro
245 250 255

CTC AGG AAG TAC ATG CTG GTA TTC AAG GGG AAG AGG TAC CTG ACA GGG 1473
Leu Arg Lys Tyr Met Leu Val Phe Lys Gly Lys Arg Tyr Leu Thr Gly
260 265 270

ATA GGA TCA GAC ACC AGG AAT GCC TTA TAT CAC GTC CAT AAC GGG GAG 1521
Ile Gly Ser Asp Thr Arg Asn Ala Leu Tyr His Val His Asn Gly Glu
275 280 285 290

GAC GTT GTG CTC CTC ACC ACC TGC AAG CAT GGC AAA GAC TGG CAA AAG 1569
Asp Val Val Leu Leu Thr Thr Cys Lys His Gly Lys Asp Trp Gln Lys
295 300 305

CAC AAG GAT TCT CGC TGT GAC AGA GAC AAC ACC GAG TAT GAG AAG TAT 1617
His Lys Asp Ser Arg Cys Asp Arg Asp Asn Thr Glu Tyr Glu Lys Tyr
310 315 320

GAT TAT CGG GAA ATG CTG CAC AAT GCC ACT TTC TGT CTG GTT CCT CGT 1665
Asp Tyr Arg Glu Met Leu His Asn Ala Thr Phe Cys Leu Val Pro Arg
325 330 335

GGT CGC AGG CTT GGG TCC TTC AGA TTC CTG GAG GCT TTG CAG GCT GCC 1713
Gly Arg Arg Leu Gly Ser Phe Arg Phe Leu Glu Ala Leu Gln Ala Ala
340 345 350

TGC GTC CCT GTG ATG CTC AGC AAT GGA TGG GAG TTG CCA TTC TCT GAA 1761
Cys Val Pro Val Met Leu Ser Asn Gly Trp Glu Leu Pro Phe Ser Glu
355 360 365 370

GTG ATT AAT TGG AAC CAA GCT GCC GTC ATA GGC GAT GAG AGA TTG TTA 1809
Val Ile Asn Trp Asn Gln Ala Ala Val Ile Gly Asp Glu Arg Leu Leu
375 380 385

TTA CAG ATT CCT TCT ACA ATC AGG TCT ATT CAT CAG GAT AAA ATC CTA 1857
Leu Gln Ile Pro Ser Thr Ile Arg Ser Ile His Gln Asp Lys Ile Leu

390 395 400

GCA CTT AGA CAG CAG ACA CAA TTC TTG TGG GAG GCT TAT TTT TCT TCA 1905
Ala Leu Arg Gln Gln Thr Gln Phe Leu Trp Glu Ala Tyr Phe Ser Ser
405 410 415

GTT GAG AAG ATT GTA TTA ACT ACA CTA GAG ATT ATT CAG GAC AGA ATA 1953
Val Glu Lys Ile Val Leu Thr Thr Leu Glu Ile Ile Gln Asp Arg Ile
420 425 430

TTC AAG CAC ATA TCA CGT AAC AGT TTA ATA TGG AAC AAA CAT CCT GGA 2001
Phe Lys His Ile Ser Arg Asn Ser Leu Ile Trp Asn Lys His Pro Gly
435 440 445 450

GGA TTG TTC GTA CTA CCA CAG TAT TCA TCT TAT CTG GGA GAT TTT CCT 2049
Gly Leu Phe Val Leu Pro Gln Tyr Ser Ser Tyr Leu Gly Asp Phe Pro
455 460 465

TAC TAC TAT GCT AAT TTA GGT TTA AAG CCC CCC TCC AAA TTC ACT GCA 2097
Tyr Tyr Tyr Ala Asn Leu Gly Leu Lys Pro Pro Ser Lys Phe Thr Ala
470 475 480

GTC ATC CAT GCG GTG ACC CCC CTG GTC TCT CAG TCC CAG CCA GTG TTG 2145
Val Ile His Ala Val Thr Pro Leu Val Ser Gln Ser Gln Pro Val Leu
485 490 495

AAG CTT CTC GTG GCT GCA GCC AAG TCC CAG TAC TGT GCC CAG ATC ATA 2193
Lys Leu Leu Val Ala Ala Ala Lys Ser Gln Tyr Cys Ala Gln Ile Ile
500 505 510

GTT CTA TGG AAT TGT GAC AAG CCC CTA CCA GCC AAA CAC CGC TGG CCT 2241
Val Leu Trp Asn Cys Asp Lys Pro Leu Pro Ala Lys His Arg Trp Pro
515 520 525 530

GCC ACT GCT GTG CCT GTC GTC GTC ATT GAA GGA GAG AGC AAG GTT ATG 2289
Ala Thr Ala Val Pro Val Val Val Ile Glu Gly Glu Ser Lys Val Met
535 540 545

AGC AGC CGT TTT CTG CCC TAC GAC AAC ATC ATC ACA GAC GCC GTG CTC 2337
Ser Ser Arg Phe Leu Pro Tyr Asp Asn Ile Ile Thr Asp Ala Val Leu
550 555 560

AGC CTT GAC GAG GAC ACG GTG CTT TCA ACA ACA GAG GTG GAT TTC GCC 2385
Ser Leu Asp Glu Asp Thr Val Leu Ser Thr Thr Glu Val Asp Phe Ala
565 570 575

TTC ACA GTG TGG CAG AGC TTC CCT GAG AGG ATT GTG GGG TAC CCC GCG 2433
Phe Thr Val Trp Gln Ser Phe Pro Glu Arg Ile Val Gly Tyr Pro Ala
580 585 590

CGC AGC CAC TTC TGG GAT AAC TCT AAG GAG CGG TGG GGA TAC ACA TCA 2481
Arg Ser His Phe Trp Asp Asn Ser Lys Glu Arg Trp Gly Tyr Thr Ser
595 600 605 610

AAG TGG ACG AAC GAC TAC TCC ATG GTG TTG ACA GGA GCT GCT ATT TAC 2529
Lys Trp Thr Asn Asp Tyr Ser Met Val Leu Thr Gly Ala Ala Ile Tyr
615 620 625

CAC AAA TAT TAT CAC TAC CTA TAC TCC CAT TAC CTG CCA GCC AGC CTG 2577
His Lys Tyr Tyr His Tyr Leu Tyr Ser His Tyr Leu Pro Ala Ser Leu

630 635 640

AAG AAC ATG GTG GAC CAA TTG GCC AAT TGT GAG GAC ATT CTC ATG AAC 2625
Lys Asn Met Val Asp Gln Leu Ala Asn Cys Glu Asp Ile Leu Met Asn
645 650 655

TTC CTG GTG TCT GCT GTG ACA AAA TTG CCT CCA ATC AAA GTG ACC CAG 2673
Phe Leu Val Ser Ala Val Thr Lys Leu Pro Pro Ile Lys Val Thr Gln
660 665 670

AAG AAG CAG TAT AAG GAG ACA ATG ATG GGA CAG ACT TCT CGG GCT TCC 2721
Lys Lys Gln Tyr Lys Glu Thr Met Met Gly Gln Thr Ser Arg Ala Ser
675 680 685 690

CGT TGG GCT GAC CCT GAC CAC TTT GCC CAG CGA CAG AGC TGC ATG AAT 2769
Arg Trp Ala Asp Pro Asp His Phe Ala Gln Arg Gln Ser Cys Met Asn
695 700 705

ACG TTT GCC AGC TGG TTT GGC TAC ATG CCG CTG ATC CAC TCT CAG ATG 2817
Thr Phe Ala Ser Trp Phe Gly Tyr Met Pro Leu Ile His Ser Gln Met
710 715 720

AGG CTC GAC CCC GTC CTC TTT AAA GAC CAG GTC TCT ATT TTG AGG AAG 2865
Arg Leu Asp Pro Val Leu Phe Lys Asp Gln Val Ser Ile Leu Arg Lys
725 730 735

AAA TAC CGA GAC ATT GAG CGA CTT TGAGGAATCC GGCTGAGTGG GGGAGGGGAA 2919
Lys Tyr Arg Asp Ile Glu Arg Leu
740 745

GCAAGAAGGG ATGGGGGTCA AGCTGCTCTC TCTTCCCAGT GCAGATCCAC TCATCAGCAG 2979

AGCCAGATTG TGCCAACTAT CCAAAAACTT AGATGAGCAG AATGACAAAA AAAAAAAAGG 3039

CCAATGAGAA CTCAACTCCT GGCTCCTGGG ACTGCACCAG ACTGCTCCAA ACTCACCTCA 3099

CTGGCTTCTG TGTCCCAAGA CTAGGTTGGT ACAGTTTAAT TATGGAACAT TAAATAATTA 3159

TTTTTGAAAA AAAAAAAAAA AAAA 3183

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 746 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gln Ala Lys Lys Arg Tyr Phe Ile Leu Leu Ser Ala Gly Ser Cys
1 5 10 15

Leu Ala Leu Leu Phe Tyr Phe Gly Gly Leu Gln Phe Arg Ala Ser Arg
20 25 30

Ser His Ser Arg Arg Glu Glu His Ser Gly Arg Asn Gly Leu His His
35 40 45

Pro Ser Pro Asp His Phe Trp Pro Arg Phe Pro Glu Pro Leu Arg Pro
50 55 60

Phe Val Pro Trp Asp Gln Leu Glu Asn Glu Asp Ser Ser Val His Ile
65 70 75 80

Ser Pro Arg Gln Lys Arg Asp Ala Asn Ser Ser Ile Tyr Lys Gly Lys
85 90 95

Lys Cys Arg Met Glu Ser Cys Phe Asp Phe Thr Leu Cys Lys Lys Asn
100 105 110

Gly Phe Lys Val Tyr Val Tyr Pro Gln Gln Lys Gly Glu Lys Ile Ala
115 120 125

Glu Ser Tyr Gln Asn Ile Leu Ala Ala Ile Glu Gly Ser Arg Phe Tyr
130 135 140

Thr Ser Asp Pro Ser Gln Ala Cys Leu Phe Val Leu Ser Leu Asp Thr
145 150 155 160

Leu Asp Arg Asp Gln Leu Ser Pro Gln Tyr Val His Asn Leu Arg Ser
165 170 175

Lys Val Gln Ser Leu His Leu Trp Asn Asn Gly Arg Asn His Leu Ile
180 185 190

Phe Asn Leu Tyr Ser Gly Thr Trp Pro Asp Tyr Thr Glu Asp Val Gly
195 200 205

Phe Asp Ile Gly Gln Ala Met Leu Ala Lys Ala Ser Ile Ser Thr Glu
210 215 220

Asn Phe Arg Pro Asn Phe Asp Val Ser Ile Pro Leu Phe Ser Lys Asp
225 230 235 240

His Pro Arg Thr Gly Gly Glu Arg Gly Phe Leu Lys Phe Asn Thr Ile
245 250 255

Pro Pro Leu Arg Lys Tyr Met Leu Val Phe Lys Gly Lys Arg Tyr Leu
260 265 270

Thr Gly Ile Gly Ser Asp Thr Arg Asn Ala Leu Tyr His Val His Asn
275 280 285

Gly Glu Asp Val Val Leu Leu Thr Thr Cys Lys His Gly Lys Asp Trp
290 295 300

Gln Lys His Lys Asp Ser Arg Cys Asp Arg Asp Asn Thr Glu Tyr Glu
305 310 315 320

Lys Tyr Asp Tyr Arg Glu Met Leu His Asn Ala Thr Phe Cys Leu Val
325 330 335

Pro Arg Gly Arg Arg Leu Gly Ser Phe Arg Phe Leu Glu Ala Leu Gln
340 345 350

Ala Ala Cys Val Pro Val Met Leu Ser Asn Gly Trp Glu Leu Pro Phe
355 360 365

Ser Glu Val Ile Asn Trp Asn Gln Ala Ala Val Ile Gly Asp Glu Arg
370 375 380

Leu Leu Leu Gln Ile Pro Ser Thr Ile Arg Ser Ile His Gln Asp Lys
385 390 395 400

Ile Leu Ala Leu Arg Gln Gln Thr Gln Phe Leu Trp Glu Ala Tyr Phe
405 410 415

Ser Ser Val Glu Lys Ile Val Leu Thr Thr Leu Glu Ile Ile Gln Asp
420 425 430

Arg Ile Phe Lys His Ile Ser Arg Asn Ser Leu Ile Trp Asn Lys His
435 440 445

Pro Gly Gly Leu Phe Val Leu Pro Gln Tyr Ser Ser Tyr Leu Gly Asp
450 455 460

Phe Pro Tyr Tyr Tyr Ala Asn Leu Gly Leu Lys Pro Pro Ser Lys Phe
465 470 475 480

Thr Ala Val Ile His Ala Val Thr Pro Leu Val Ser Gln Ser Gln Pro
485 490 495

Val Leu Lys Leu Leu Val Ala Ala Ala Lys Ser Gln Tyr Cys Ala Gln
500 505 510

Ile Ile Val Leu Trp Asn Cys Asp Lys Pro Leu Pro Ala Lys His Arg
515 520 525

Trp Pro Ala Thr Ala Val Pro Val Val Val Ile Glu Gly Glu Ser Lys
530 535 540

Val Met Ser Ser Arg Phe Leu Pro Tyr Asp Asn Ile Ile Thr Asp Ala
545 550 555 560

Val Leu Ser Leu Asp Glu Asp Thr Val Leu Ser Thr Thr Glu Val Asp
565 570 575

Phe Ala Phe Thr Val Trp Gln Ser Phe Pro Glu Arg Ile Val Gly Tyr
580 585 590

Pro Ala Arg Ser His Phe Trp Asp Asn Ser Lys Glu Arg Trp Gly Tyr
595 600 605

Thr Ser Lys Trp Thr Asn Asp Tyr Ser Met Val Leu Thr Gly Ala Ala
610 615 620

Ile Tyr His Lys Tyr Tyr His Tyr Leu Tyr Ser His Tyr Leu Pro Ala
625 630 635 640

Ser Leu Lys Asn Met Val Asp Gln Leu Ala Asn Cys Glu Asp Ile Leu
645 650 655

Met Asn Phe Leu Val Ser Ala Val Thr Lys Leu Pro Pro Ile Lys Val
660 665 670

Thr Gln Lys Lys Gln Tyr Lys Glu Thr Met Met Gly Gln Thr Ser Arg
675 680 685

```
Ala Ser Arg Trp Ala Asp Pro Asp His Phe Ala Gln Arg Gln Ser Cys
    690                 695                 700

Met Asn Thr Phe Ala Ser Trp Phe Gly Tyr Met Pro Leu Ile His Ser
705                 710                 715                 720

Gln Met Arg Leu Asp Pro Val Leu Phe Lys Asp Gln Val Ser Ile Leu
                725                 730                 735

Arg Lys Lys Tyr Arg Asp Ile Glu Arg Leu
            740                 745
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 335..2488

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CTGTCTGAGC ATTTCACTGC GGAGCCTGAG CGCGCCTGCC TGGGAAAACA CTGCAGCGGT      60

GCTCGGACTC CTCCTGTCCA GCAGGAGGCG CGGCCCGGCA GCTCCCGCAT GCGCAGTGCG     120

CTCGGTGTCA GACGGCCCGG ATCCCGGTTA CCGGCCCCTC GCTCGCTGCT CGCCAGCCCA     180

GACTCGGCCC TGGCAGTGGC GGCTGGCGAT TCGGACCGAT CCGACCTGGG CGGAGGTGGC     240

CCGCGCCCCG CGGCATGAGC CGGTGACCAA GCTCGGGGCC GAGCGGGAGG CAGCCGTGGC     300

CGAGGAGTGT GAGGAAGAGG CTGTCTGTGT CATT ATG TGT GCG TCG GTC AAG        352
                                      Met Cys Ala Ser Val Lys
                                      1               5

TAT AAT ATC CGG GGT CCT GCC CTC ATC CCA AGA ATG AAG ACC AAG CAC      400
Tyr Asn Ile Arg Gly Pro Ala Leu Ile Pro Arg Met Lys Thr Lys His
            10                  15                  20

CGA ATC TAC TAT ATC ACC CTC TTC TCC ATT GTC CTC CTG GGC CTC ATT      448
Arg Ile Tyr Tyr Ile Thr Leu Phe Ser Ile Val Leu Leu Gly Leu Ile
        25                  30                  35

GCC ACT GGC ATG TTT CAG TTT TGG CCC CAT TCT ATC GAG TCC TCA AAT      496
Ala Thr Gly Met Phe Gln Phe Trp Pro His Ser Ile Glu Ser Ser Asn
    40                  45                  50

GAC TGG AAT GTA GAG AAG CGC AGC ATC CGT GAT GTG CCG GTT GTT AGG      544
Asp Trp Asn Val Glu Lys Arg Ser Ile Arg Asp Val Pro Val Val Arg
55                  60                  65                  70
```

```
CTG CCA GCC GAC AGT CCC ATC CCA GAG CGG GGG GAT CTC AGT TGC AGA      592
Leu Pro Ala Asp Ser Pro Ile Pro Glu Arg Gly Asp Leu Ser Cys Arg
                 75                  80                  85

ATG CAC ACG TGT TTT GAT GTC TAT CGC TGT GGC TTC AAC CCA AAG AAC      640
Met His Thr Cys Phe Asp Val Tyr Arg Cys Gly Phe Asn Pro Lys Asn
             90                  95                 100

AAA ATC AAG GTG TAT ATC TAT GCT CTG AAA AAG TAC GTG GAT GAC TTT      688
Lys Ile Lys Val Tyr Ile Tyr Ala Leu Lys Lys Tyr Val Asp Asp Phe
        105                 110                 115

GGC GTC TCT GTC AGC AAC ACC ATC TCC CGG GAG TAT AAT GAA CTG CTC      736
Gly Val Ser Val Ser Asn Thr Ile Ser Arg Glu Tyr Asn Glu Leu Leu
    120                 125                 130

ATG GCC ATC TCA GAC AGT GAC TAC TAC ACT GAT GAC ATC AAC CGG GCC      784
Met Ala Ile Ser Asp Ser Asp Tyr Tyr Thr Asp Asp Ile Asn Arg Ala
135                 140                 145                 150

TGT CTG TTT GTT CCC TCC ATC GAT GTG CTT AAC CAG AAC ACA CTG CGC      832
Cys Leu Phe Val Pro Ser Ile Asp Val Leu Asn Gln Asn Thr Leu Arg
                155                 160                 165

ATC AAG GAG ACA GCA CAA GCG ATG GCC CAG CTC TCT AGG TGG GAT CGA      880
Ile Lys Glu Thr Ala Gln Ala Met Ala Gln Leu Ser Arg Trp Asp Arg
            170                 175                 180

GGT ACG AAT CAC CTG TTG TTC AAC ATG TTG CCT GGA GGT CCC CCA GAT      928
Gly Thr Asn His Leu Leu Phe Asn Met Leu Pro Gly Gly Pro Pro Asp
        185                 190                 195

TAT AAC ACA GCC CTG GAT GTC CCC AGA GAC AGG GCC CTG TTG GCT GGT      976
Tyr Asn Thr Ala Leu Asp Val Pro Arg Asp Arg Ala Leu Leu Ala Gly
    200                 205                 210

GGC GGC TTT TCT ACG TGG ACT TAC CGG CAA GGC TAC GAT GTC AGC ATT     1024
Gly Gly Phe Ser Thr Trp Thr Tyr Arg Gln Gly Tyr Asp Val Ser Ile
215                 220                 225                 230

CCT GTC TAT AGT CCA CTG TCA GCT GAG GTG GAT CTT CCA GAG AAA GGA     1072
Pro Val Tyr Ser Pro Leu Ser Ala Glu Val Asp Leu Pro Glu Lys Gly
                235                 240                 245

CCA GGT CCA CGG CAA TAC TTC CTC CTG TCA TCT CAG GTG GGT CTC CAT     1120
Pro Gly Pro Arg Gln Tyr Phe Leu Leu Ser Ser Gln Val Gly Leu His
            250                 255                 260

CCT GAG TAC AGA GAG GAC CTA GAA GCC CTC CAG GTC AAA CAT GGA GAG     1168
Pro Glu Tyr Arg Glu Asp Leu Glu Ala Leu Gln Val Lys His Gly Glu
        265                 270                 275

TCA GTG TTA GTA CTC GAT AAA TGC ACC AAC CTC TCA GAG GGT GTC CTT     1216
Ser Val Leu Val Leu Asp Lys Cys Thr Asn Leu Ser Glu Gly Val Leu
    280                 285                 290

TCT GTC CGT AAG CGC TGC CAC AAG CAC CAG GTC TTC GAT TAC CCA CAG     1264
Ser Val Arg Lys Arg Cys His Lys His Gln Val Phe Asp Tyr Pro Gln
295                 300                 305                 310
```

GTG CTA CAG GAG GCT ACT TTC TGT GTG GTT CTT CGT GGA GCT CGG CTG
Val Leu Gln Glu Ala Thr Phe Cys Val Val Leu Arg Gly Ala Arg Leu 1312
315 320 325

GGC CAG GCA GTA TTG AGC GAT GTG TTA CAA GCT GGC TGT GTC CCG GTT
Gly Gln Ala Val Leu Ser Asp Val Leu Gln Ala Gly Cys Val Pro Val 1360
330 335 340

GTC ATT GCA GAC TCC TAT ATT TTG CCT TTC TCT GAA GTT CTT GAC TGG
Val Ile Ala Asp Ser Tyr Ile Leu Pro Phe Ser Glu Val Leu Asp Trp 1408
345 350 355

AAG AGA GCA TCT GTG GTT GTA CCA GAA GAA AAG ATG TCA GAT GTG TAC
Lys Arg Ala Ser Val Val Val Pro Glu Glu Lys Met Ser Asp Val Tyr 1456
360 365 370

AGT ATT TTG CAG AGC ATC CCC CAA AGA CAG ATT GAA GAA ATG CAG AGA
Ser Ile Leu Gln Ser Ile Pro Gln Arg Gln Ile Glu Glu Met Gln Arg 1504
375 380 385 390

CAG GCC CGG TGG TTC TGG GAA GCG TAC TTC CAG TCA ATT AAA GCC ATT
Gln Ala Arg Trp Phe Trp Glu Ala Tyr Phe Gln Ser Ile Lys Ala Ile 1552
395 400 405

GCC CTG GCC ACC CTG CAG ATT ATC AAT GAC CGG ATC TAT CCA TAT GCT
Ala Leu Ala Thr Leu Gln Ile Ile Asn Asp Arg Ile Tyr Pro Tyr Ala 1600
410 415 420

GCC ATC TCC TAT GAA GAA TGG AAT GAC CCT CCT GCT GTG AAG TGG GGC
Ala Ile Ser Tyr Glu Glu Trp Asn Asp Pro Pro Ala Val Lys Trp Gly 1648
425 430 435

AGC GTG AGC AAT CCA CTC TTC CTC CCG CTG ATC CCA CCA CAG TCT CAA
Ser Val Ser Asn Pro Leu Phe Leu Pro Leu Ile Pro Pro Gln Ser Gln 1696
440 445 450

GGG TTC ACC GCC ATA GTC CTC ACC TAC GAC CGA GTA GAG AGC CTC TTC
Gly Phe Thr Ala Ile Val Leu Thr Tyr Asp Arg Val Glu Ser Leu Phe 1744
455 460 465 470

CGG GTC ATC ACT GAA GTG TCC AAG GTG CCC AGT CTA TCC AAA CTA CTT
Arg Val Ile Thr Glu Val Ser Lys Val Pro Ser Leu Ser Lys Leu Leu 1792
475 480 485

GTC GTC TGG AAT AAT CAG AAT AAA AAC CCT CCA GAA GAT TCT CTC TGG
Val Val Trp Asn Asn Gln Asn Lys Asn Pro Pro Glu Asp Ser Leu Trp 1840
490 495 500

CCC AAA ATC CGG GTT CCA TTA AAA GTT GTG AGG ACT GCT GAA AAC AAG
Pro Lys Ile Arg Val Pro Leu Lys Val Val Arg Thr Ala Glu Asn Lys 1888
505 510 515

TTA AGT AAC CGT TTC TTC CCT TAT GAT GAA ATC GAG ACA GAA GCT GTT
Leu Ser Asn Arg Phe Phe Pro Tyr Asp Glu Ile Glu Thr Glu Ala Val 1936
520 525 530

CTG GCC ATT GAT GAT GAT ATC ATT ATG CTG ACC TCT GAC GAG CTG CAA
Leu Ala Ile Asp Asp Asp Ile Ile Met Leu Thr Ser Asp Glu Leu Gln 1984
535 540 545 550

TTT GGT TAT GAG GTC TGG CGG GAA TTT CCT GAC CGG TTG GTG GGT TAC
Phe Gly Tyr Glu Val Trp Arg Glu Phe Pro Asp Arg Leu Val Gly Tyr 2032
555 560 565

CCG GGT CGT CTG CAT CTC TGG GAC CAT GAG ATG AAT AAG TGG AAG TAT
Pro Gly Arg Leu His Leu Trp Asp His Glu Met Asn Lys Trp Lys Tyr 2080
570 575 580

GAG TCT GAG TGG ACG AAT GAA GTG TCC ATG GTG CTC ACT GGG GCA GCT
Glu Ser Glu Trp Thr Asn Glu Val Ser Met Val Leu Thr Gly Ala Ala 2128
585 590 595

TTT TAT CAC AAG TAT TTT AAT TAC CTG TAT ACC TAC AAA ATG CCT GGG
Phe Tyr His Lys Tyr Phe Asn Tyr Leu Tyr Thr Tyr Lys Met Pro Gly 2176
600 605 610

GAT ATC AAG AAC TGG GTA GAT GCT CAT ATG AAC TGT GAA GAT ATT GCC
Asp Ile Lys Asn Trp Val Asp Ala His Met Asn Cys Glu Asp Ile Ala 2224
615 620 625 630

ATG AAC TTC CTG GTG GCC AAC GTC ACG GGA AAA GCA GTT ATC AAG GTA
Met Asn Phe Leu Val Ala Asn Val Thr Gly Lys Ala Val Ile Lys Val 2272
635 640 645

ACC CCA CGA AAG AAA TTC AAG TGT CCT GAG TGC ACA GCC ATA GAT GGG
Thr Pro Arg Lys Lys Phe Lys Cys Pro Glu Cys Thr Ala Ile Asp Gly 2320
650 655 660

CTT TCA CTA GAC CAA ACA CAC ATG GTG GAG AGG TCA GAG TGC ATC AAC
Leu Ser Leu Asp Gln Thr His Met Val Glu Arg Ser Glu Cys Ile Asn 2368
665 670 675

AAG TTT GCT TCA GTC TTC GGG ACC ATG CCT CTC AAG GTG GTG GAA CAC
Lys Phe Ala Ser Val Phe Gly Thr Met Pro Leu Lys Val Val Glu His 2416
680 685 690

CGA GCT GAC CCT GTC CTG TAC AAA GAT GAC TTT CCT GAG AAG CTG AAG
Arg Ala Asp Pro Val Leu Tyr Lys Asp Asp Phe Pro Glu Lys Leu Lys 2464
695 700 705 710

AGC TTC CCC AAC ATT GGC AGC TTA TGAAACGTGT CATTGGTGGA GGTCTGAATG
Ser Phe Pro Asn Ile Gly Ser Leu 2518
715

TGAGGCTGGG ACAGAGGGAG AGAACAAGGC CTCCCAGCAC TCTGATGTCA GAGTAGTAGG 2578

TTAAGGGTGG AAGGTTGACC TACTTGGATC TTGGCATGCA CCCACCTAAC CCACTTTCTC 2638

AAGAACAAGA ACCTAGAATG AATATCCAAG CACCTCGAGC TATGCAACCT CTGTTCTTGT 2698

ATTTCTTATG ATCTCTGATG GGTTCTTCTC GAAAATGCCA AGTGGAAGAC TTTGTGGCAT 2758

GCTCCAGATT TAAATCCAGC TGAGGCTCCC TTTGTTTTCA GTTCCATGTA ACAATCTGGA 2818

AGGAAACTTC ACGGACAGGA AGACTGCTGG AGAAGAGAAG CGTGTTAGCC CATTTGAGGT 2878

CTGGGGAATC ATGTAAAGGG TACCCAGACC TCACTTTTAG TTATTTACAT CAATGAGTTC 2938

TTTCAGGGAA CCAAACCCAG AATTCGGTGC AAAAGCCAAA CATCTTGGTG GGATTTGATA 2998

AATGCCTTGG GACCTGGAGT GCTGGGCTTG TGCACAGGAA GAGCACCAGC CGCTGAGTCA 3058

GGATCCTGTC AGTTCCATGA GCTATTCCTC TTTGGTTTGG CTTTTTGATA TGATTAAAAT 3118

TATTTTTTAT TCCTTTTAAA AAAAAAAAAA AAAAAAAAAA TTCGTCGTGC TTAAACA 3175

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 718 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Cys Ala Ser Val Lys Tyr Asn Ile Arg Gly Pro Ala Leu Ile Pro
1 5 10 15

Arg Met Lys Thr Lys His Arg Ile Tyr Tyr Ile Thr Leu Phe Ser Ile
20 25 30

Val Leu Leu Gly Leu Ile Ala Thr Gly Met Phe Gln Phe Trp Pro His
35 40 45

Ser Ile Glu Ser Ser Asn Asp Trp Asn Val Glu Lys Arg Ser Ile Arg
50 55 60

Asp Val Pro Val Val Arg Leu Pro Ala Asp Ser Pro Ile Pro Glu Arg
65 70 75 80

Gly Asp Leu Ser Cys Arg Met His Thr Cys Phe Asp Val Tyr Arg Cys
85 90 95

Gly Phe Asn Pro Lys Asn Lys Ile Lys Val Tyr Ile Tyr Ala Leu Lys
100 105 110

Lys Tyr Val Asp Asp Phe Gly Val Ser Val Ser Asn Thr Ile Ser Arg
115 120 125

Glu Tyr Asn Glu Leu Leu Met Ala Ile Ser Asp Ser Asp Tyr Tyr Thr
130 135 140

Asp Asp Ile Asn Arg Ala Cys Leu Phe Val Pro Ser Ile Asp Val Leu
145 150 155 160

Asn Gln Asn Thr Leu Arg Ile Lys Glu Thr Ala Gln Ala Met Ala Gln
165 170 175

Leu Ser Arg Trp Asp Arg Gly Thr Asn His Leu Leu Phe Asn Met Leu
180 185 190

Pro Gly Gly Pro Pro Asp Tyr Asn Thr Ala Leu Asp Val Pro Arg Asp
195 200 205

Arg Ala Leu Leu Ala Gly Gly Gly Phe Ser Thr Trp Thr Tyr Arg Gln
210 215 220

Gly Tyr Asp Val Ser Ile Pro Val Tyr Ser Pro Leu Ser Ala Glu Val
225 230 235 240

Asp Leu Pro Glu Lys Gly Pro Gly Pro Arg Gln Tyr Phe Leu Leu Ser
245 250 255

Ser Gln Val Gly Leu His Pro Glu Tyr Arg Glu Asp Leu Glu Ala Leu
260 265 270

Gln Val Lys His Gly Glu Ser Val Leu Val Leu Asp Lys Cys Thr Asn
275 280 285

Leu Ser Glu Gly Val Leu Ser Val Arg Lys Arg Cys His Lys His Gln
290 295 300

Val Phe Asp Tyr Pro Gln Val Leu Gln Glu Ala Thr Phe Cys Val Val
305 310 315 320

Leu Arg Gly Ala Arg Leu Gly Gln Ala Val Leu Ser Asp Val Leu Gln
325 330 335

Ala Gly Cys Val Pro Val Val Ile Ala Asp Ser Tyr Ile Leu Pro Phe
340 345 350

Ser Glu Val Leu Asp Trp Lys Arg Ala Ser Val Val Val Pro Glu Glu
355 360 365

Lys Met Ser Asp Val Tyr Ser Ile Leu Gln Ser Ile Pro Gln Arg Gln
370 375 380

Ile Glu Glu Met Gln Arg Gln Ala Arg Trp Phe Trp Glu Ala Tyr Phe
385 390 395 400

Gln Ser Ile Lys Ala Ile Ala Leu Ala Thr Leu Gln Ile Ile Asn Asp
405 410 415

Arg Ile Tyr Pro Tyr Ala Ala Ile Ser Tyr Glu Glu Trp Asn Asp Pro
420 425 430

Pro Ala Val Lys Trp Gly Ser Val Ser Asn Pro Leu Phe Leu Pro Leu
435 440 445

Ile Pro Pro Gln Ser Gln Gly Phe Thr Ala Ile Val Leu Thr Tyr Asp
450 455 460

Arg Val Glu Ser Leu Phe Arg Val Ile Thr Glu Val Ser Lys Val Pro
465 470 475 480

Ser Leu Ser Lys Leu Leu Val Val Trp Asn Asn Gln Asn Lys Asn Pro
485 490 495

Pro Glu Asp Ser Leu Trp Pro Lys Ile Arg Val Pro Leu Lys Val Val
500 505 510

Arg Thr Ala Glu Asn Lys Leu Ser Asn Arg Phe Phe Pro Tyr Asp Glu
515 520 525

Ile Glu Thr Glu Ala Val Leu Ala Ile Asp Asp Asp Ile Ile Met Leu
530 535 540

Thr Ser Asp Glu Leu Gln Phe Gly Tyr Glu Val Trp Arg Glu Phe Pro
545 550 555 560

Asp Arg Leu Val Gly Tyr Pro Gly Arg Leu His Leu Trp Asp His Glu
565 570 575

Met Asn Lys Trp Lys Tyr Glu Ser Glu Trp Thr Asn Glu Val Ser Met
580 585 590

Val Leu Thr Gly Ala Ala Phe Tyr His Lys Tyr Phe Asn Tyr Leu Tyr
595 600 605

Thr Tyr Lys Met Pro Gly Asp Ile Lys Asn Trp Val Asp Ala His Met
610 615 620

Asn Cys Glu Asp Ile Ala Met Asn Phe Leu Val Ala Asn Val Thr Gly
625 630 635 640

Lys Ala Val Ile Lys Val Thr Pro Arg Lys Lys Phe Lys Cys Pro Glu
645 650 655

Cys Thr Ala Ile Asp Gly Leu Ser Leu Asp Gln Thr His Met Val Glu
660 665 670

Arg Ser Glu Cys Ile Asn Lys Phe Ala Ser Val Phe Gly Thr Met Pro
675 680 685

Leu Lys Val Val Glu His Arg Ala Asp Pro Val Leu Tyr Lys Asp Asp
690 695 700

Phe Pro Glu Lys Leu Lys Ser Phe Pro Asn Ile Gly Ser Leu
705 710 715

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3642 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 263..2542

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAATAAAACA AAATAACCAA AATCGAAAAT GCGCGACAAC TAGTGTAAAT TCGTGTTGTT 60

ATTGTTGTTG TCGGAACCCC CGCCTCACTC TCTCTCTCTC TGCCGCACAC ACACCATCAC 120

ACACACGCAC GCACGCTCTC CCACATACAC AAACACACGC ACACCCATAC AGCATCCCCG 180

GAAAATAAAT AAAAGTAAAT AATAATAATA TAAATCAAAA TTATTGCGGA TCTTAAGGCT 240

```
TTGGATACAC TAAAACCACA AA ATG CAG GCC AAA AAG CGT TAT ATT TTG GTC     292
                         Met Gln Ala Lys Lys Arg Tyr Ile Leu Val
                          1               5                  10

TTC GTT TCC TGC GCT TTT TTG GCC TAC GCC TAT TTC GGT GGC TAT CGC     340
Phe Val Ser Cys Ala Phe Leu Ala Tyr Ala Tyr Phe Gly Gly Tyr Arg
                 15                  20                  25

CTG AAA GTC TCA CCA TTG AGA CCC CGT AGA GCC CAG CAC GAA TCG GCC     388
Leu Lys Val Ser Pro Leu Arg Pro Arg Arg Ala Gln His Glu Ser Ala
             30                  35                  40

AAG GAT GGT GGA GTT CAA CCC CAC GAG CAG TTG CCC AGC TTC CTG GGC     436
Lys Asp Gly Gly Val Gln Pro His Glu Gln Leu Pro Ser Phe Leu Gly
         45                  50                  55

GCC CAC GAT ATG CAG GAA CTC CAA CTG CTG CAG AGC AAT CAA TCG AAG     484
Ala His Asp Met Gln Glu Leu Gln Leu Leu Gln Ser Asn Gln Ser Lys
     60                  65                  70

AGT TTG GAT AGC TCC AAG CAC CTG GTT ACC CGT AAA CCC GAC TGC CGC     532
Ser Leu Asp Ser Ser Lys His Leu Val Thr Arg Lys Pro Asp Cys Arg
 75                  80                  85                  90

ATG GAG ACC TGT TTC GAT TTT ACC CGC TGT TAT GAT CGC TTT TTG GTC     580
Met Glu Thr Cys Phe Asp Phe Thr Arg Cys Tyr Asp Arg Phe Leu Val
                 95                 100                 105

TAT ATC TAT CCA CCG GAA CCA CTT AAC TCA CTG GGC GCT GCC CCG CCC     628
Tyr Ile Tyr Pro Pro Glu Pro Leu Asn Ser Leu Gly Ala Ala Pro Pro
            110                 115                 120

ACC TCG GCC AAC TAT CAA AAG ATA CTC ACT GCC ATC CAG GAA TCG AGA     676
Thr Ser Ala Asn Tyr Gln Lys Ile Leu Thr Ala Ile Gln Glu Ser Arg
        125                 130                 135

TAT TAT ACC AGT GAT CCC ACG GCC GCC TGT CTC TTT GTG CTC GGT ATC     724
Tyr Tyr Thr Ser Asp Pro Thr Ala Ala Cys Leu Phe Val Leu Gly Ile
    140                 145                 150

GAT ACC CTG GAT AGA GAT TCC CTA TCC GAG GAT TAT GTT CGA AAT GTG     772
Asp Thr Leu Asp Arg Asp Ser Leu Ser Glu Asp Tyr Val Arg Asn Val
155                 160                 165                 170

CCA TCC AGA TTG GCA AGA TTG CCG TAT TGG AAC AAT GGC AGG AAC CAC     820
Pro Ser Arg Leu Ala Arg Leu Pro Tyr Trp Asn Asn Gly Arg Asn His
                175                 180                 185

ATT ATC TTC AAT CTG TAC TCG GGC ACT TGG CCG GAT TAC GCG GAG AAC     868
Ile Ile Phe Asn Leu Tyr Ser Gly Thr Trp Pro Asp Tyr Ala Glu Asn
            190                 195                 200

TCA TTG GGT TTC GAT GCG GGT GAG GCC ATT TTG GCC AAG GCC AGC ATG     916
Ser Leu Gly Phe Asp Ala Gly Glu Ala Ile Leu Ala Lys Ala Ser Met
        205                 210                 215

GGA GTG CTC CAA CTG CGA CAC GGT TTC GAC GTC AGC ATC CCG CTC TTC     964
Gly Val Leu Gln Leu Arg His Gly Phe Asp Val Ser Ile Pro Leu Phe
    220                 225                 230
```

CAC AAA CAA TTC CCG CTG AGG GCC GGC GCC ACA GGA ACC GTG CAG TCC 1012
His Lys Gln Phe Pro Leu Arg Ala Gly Ala Thr Gly Thr Val Gln Ser
235 240 245 250

AAC AAT TTC CCC GCC AAC AAG AAG TAT CTG CTG GCC TTC AAA GGA AAG 1060
Asn Asn Phe Pro Ala Asn Lys Lys Tyr Leu Leu Ala Phe Lys Gly Lys
255 260 265

CGA TAC GTG CAT GGC ATT GGA TCG GAG ACG CGC AAC TCG CTC TTT CAT 1108
Arg Tyr Val His Gly Ile Gly Ser Glu Thr Arg Asn Ser Leu Phe His
270 275 280

TTG CAC AAC GGA CGC GAC ATG GTC CTG GTC ACC ACC TGT CGC CAC GGG 1156
Leu His Asn Gly Arg Asp Met Val Leu Val Thr Thr Cys Arg His Gly
285 290 295

AAA AGT TGG CGG GAG CTG CAG GAT AAT CGC TGC GAC GAG GAC AAC CGG 1204
Lys Ser Trp Arg Glu Leu Gln Asp Asn Arg Cys Asp Glu Asp Asn Arg
300 305 310

GAA TAC GAC AGA TAC GAC TAC GAG ACC TTA CTG CAA AAT TCC ACT TTC 1252
Glu Tyr Asp Arg Tyr Asp Tyr Glu Thr Leu Leu Gln Asn Ser Thr Phe
315 320 325 330

TGC TTG GTG CCG CGC GGA CGT CGT TTG GGT TCC TTT AGA TTC CTC GAA 1300
Cys Leu Val Pro Arg Gly Arg Arg Leu Gly Ser Phe Arg Phe Leu Glu
335 340 345

GCC TTG CAG GCT GGT TGC ATT CCC GTA CTA TTG TCC AAT GCC TGG GTG 1348
Ala Leu Gln Ala Gly Cys Ile Pro Val Leu Leu Ser Asn Ala Trp Val
350 355 360

TTG CCG TTT GAA TCG AAA ATC GAT TGG AAG CAG GCC GCC ATT TGG GCC 1396
Leu Pro Phe Glu Ser Lys Ile Asp Trp Lys Gln Ala Ala Ile Trp Ala
365 370 375

GAT GAG CGG CTC CTG CTG CAG GTA CCC GAC ATC GTG CGC TCC ATA CCG 1444
Asp Glu Arg Leu Leu Leu Gln Val Pro Asp Ile Val Arg Ser Ile Pro
380 385 390

GCG GAA CGC ATC TTC GCC CTC CGC CAG CAG ACG CAG GTC CTG TGG GAA 1492
Ala Glu Arg Ile Phe Ala Leu Arg Gln Gln Thr Gln Val Leu Trp Glu
395 400 405 410

CGC TAC TTC GGA TCC ATT GAG AAA ATC GTC TTC ACA ACA TTC GAG ATC 1540
Arg Tyr Phe Gly Ser Ile Glu Lys Ile Val Phe Thr Thr Phe Glu Ile
415 420 425

ATA CGC GAA CGT TTG CCG GAT TAT CCG GTG CGA AGC AGC CTG GTT TGG 1588
Ile Arg Glu Arg Leu Pro Asp Tyr Pro Val Arg Ser Ser Leu Val Trp
430 435 440

AAC AGC TCG CCG GGC GCC CTG CTC ACG TTG CCC ACC TTT GCC GAC AGC 1636
Asn Ser Ser Pro Gly Ala Leu Leu Thr Leu Pro Thr Phe Ala Asp Ser
445 450 455

TCA AGG TAC ATG CCG TTT CTG CTC AAC TCG ATG GGT GCG GAG CCG CGT 1684
Ser Arg Tyr Met Pro Phe Leu Leu Asn Ser Met Gly Ala Glu Pro Arg
460 465 470

CAC AAC TAC ACG GCG GTT ATC TAT GTG CAG ATT GGC GCC GCC CTG GGA 1732
His Asn Tyr Thr Ala Val Ile Tyr Val Gln Ile Gly Ala Ala Leu Gly
475 480 485 490

CCG AAT GCA GCG CTC TAC AAG CTA GTC AGG ACC ATC ACG AAG AGC CAG 1780
Pro Asn Ala Ala Leu Tyr Lys Leu Val Arg Thr Ile Thr Lys Ser Gln
495 500 505

TTC GTG GAG CGG ATC CTT GTC CTT TGG GCA GCC GAT CGG CCG CTG CCG 1828
Phe Val Glu Arg Ile Leu Val Leu Trp Ala Ala Asp Arg Pro Leu Pro
510 515 520

TTG AAG AAA CGC TGG CCC CCC ACC AGC CAC ATA CCC TTG CAC GTG ATT 1876
Leu Lys Lys Arg Trp Pro Pro Thr Ser His Ile Pro Leu His Val Ile
525 530 535

TCC CTG GGG GGC AGC ACG AGG AGT CAG GGA GCT GGA CCC ACA AGC CAG 1924
Ser Leu Gly Gly Ser Thr Arg Ser Gln Gly Ala Gly Pro Thr Ser Gln
540 545 550

ACG ACT GAG GGA CGA CCT AGC ATA TCA CAA CGA TTC CTG CCC TAC GAT 1972
Thr Thr Glu Gly Arg Pro Ser Ile Ser Gln Arg Phe Leu Pro Tyr Asp
555 560 565 570

GAG ATT CAG ACG GAT GCC GTT CTG TCG CTG GAT GAA GAT GCC ATA CTC 2020
Glu Ile Gln Thr Asp Ala Val Leu Ser Leu Asp Glu Asp Ala Ile Leu
575 580 585

AAT ACG GAT GAA CTG GAC TTT GCC TAC ACG GTG TGG CGG GAT TTT CCA 2068
Asn Thr Asp Glu Leu Asp Phe Ala Tyr Thr Val Trp Arg Asp Phe Pro
590 595 600

GAG CGC ATC GTT GGC TAT CCG GCG AGA GCT CAC TTC TGG GAT GAT TCC 2116
Glu Arg Ile Val Gly Tyr Pro Ala Arg Ala His Phe Trp Asp Asp Ser
605 610 615

AAG AAT GCC TGG GGT TAT ACG TCC AAG TGG ACA AAC TAC TAT TCG ATT 2164
Lys Asn Ala Trp Gly Tyr Thr Ser Lys Trp Thr Asn Tyr Tyr Ser Ile
620 625 630

GTG CTA ACT GGG GCG GCA TTC TAC CAC CGC TAC TAC AAC TAC TTG TAC 2212
Val Leu Thr Gly Ala Ala Phe Tyr His Arg Tyr Tyr Asn Tyr Leu Tyr
635 640 645 650

ACC AAT TGG CTG TCG TTG CTG CTA CTT AAG ACT GTA CAG CAG TCC TCC 2260
Thr Asn Trp Leu Ser Leu Leu Leu Leu Lys Thr Val Gln Gln Ser Ser
655 660 665

AAC TGC GAG GAC ATC CTA ATG AAC CTG CTG GTC TCG CAC GTG ACC AGA 2308
Asn Cys Glu Asp Ile Leu Met Asn Leu Leu Val Ser His Val Thr Arg
670 675 680

AAG CCG CCG ATC AAG GTG ACG CAG CGC AAA GGG CTA CAA GGA TCG GGA 2356
Lys Pro Pro Ile Lys Val Thr Gln Arg Lys Gly Leu Gln Gly Ser Gly
685 690 695

GAC GGG TCG TTC GCC TGG AAC GAT CCC GAC CAC TTT ATC CAG CGC CAG 2404
Asp Gly Ser Phe Ala Trp Asn Asp Pro Asp His Phe Ile Gln Arg Gln
700 705 710

```
AGC TGC CTG AAT ACC TTT GCG GCC GTG TTT GGC TAT ATG CCG CTC ATA       2452
Ser Cys Leu Asn Thr Phe Ala Ala Val Phe Gly Tyr Met Pro Leu Ile
715                 720                 725                 730

CGT TCC AAT TTG CGC ATG GAT CCC ATG CTG TAT CGT GAT CCA GTT TCC       2500
Arg Ser Asn Leu Arg Met Asp Pro Met Leu Tyr Arg Asp Pro Val Ser
                735                 740                 745

AAT TTG CGC AAG AAA TAC CGC CAG ATC GAG TTG GTC GGC AGC               2542
Asn Leu Arg Lys Lys Tyr Arg Gln Ile Glu Leu Val Gly Ser
            750                 755                 760

TAGCGATATG TGCATCCTGG CAATAGCTAC AAGCGTTAAT AGTACTGAAT ACATCTACAC     2602

CACACACCCA CAGCTACACC AATACCAATA CCAACACCAA CATCAACACC AATAACCATA     2662

CGAATACGAA CACAGAGAAA CAGAAAAAAC ACACACACAT ATGTTTAGAT GAAATGCTGT     2722

TTTTTGACGC TGGCCAAGCG ATGCAAGAAC TATGTTAATT ATTATTAGTA GCGAAAAGAT     2782

GAAACTTAGT TAGGAGCCAA AGGGCGCTTG TAAATGTATG TATGTGTACA TCGATGTAAG     2842

CATACGGCAT CCTTTGCAAT TATATATGCA ATGTGTATAT ATATATAGTC CTTAAATACG     2902

ATTATGTGAA TTAGCCTAGT TAACTGCAAG CGAAGCAAAC GATTTTGCCA TCTTTTTTGC     2962

TAATTTCAAA TTAGCTGCGC TTAGCAAATT TTCTATATGC ATACATACAA TATGTAAATT     3022

GATATATGTC TAGGAACATA AGTAGCCGAT TTTTGTACTG TGTGAGTTGA GTTTTGTATG     3082

TGTAAATGTT TGTATGTCCT TGTTGTAACT CTCCGTGTAT AGTTAGTGAT TTATCAATGT     3142

GAACGTGTAT GCGCAAGGCA CAGTCGTCTC AATTTCAGTT TTAAATTCCA ACTTGAGGCT     3202

CCTCGAATTC CGTATAAAAA ACACATCGAC TGTATCTTTT GATTATCCGT GGCCATTATA     3262

GAACAGCAGC AGAAATTGCC ACTTAAAAGC AATACGAACT CTTTTTCGGG GCCATTCAGA     3322

AATAAGTGCA CAAATCCGCC GAGGAGAAGC AATTAACTAA TGCTAAGTTT ATGGTTTAAG     3382

TAGCTTTCCA TTTTCAATTC GTTATGTGCA ATTTTTTATT TGGACACCTC AGTTTTTTAA     3442

TTCTTATCGT GGCCATTTAC TGAAGTACGT AAATCAATTA TCAATATGCA ATCAATTTTA     3502

ACCACATGCG CATATCAAAA GAAGAGGCTT TTTTGAAGCG AAGCGTTTTT AGTTTAAGCA     3562

ACCACAATTT TATACATATA CTAGAGCTAA AGTATATTTC TATAACTATT TTTTACAACA     3622

AATATACACG ATTGCAGATG                                                  3642
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 760 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gln Ala Lys Lys Arg Tyr Ile Leu Val Phe Val Ser Cys Ala Phe
1 5 10 15

Leu Ala Tyr Ala Tyr Phe Gly Gly Tyr Arg Leu Lys Val Ser Pro Leu
20 25 30

Arg Pro Arg Arg Ala Gln His Glu Ser Ala Lys Asp Gly Gly Val Gln
35 40 45

Pro His Glu Gln Leu Pro Ser Phe Leu Gly Ala His Asp Met Gln Glu
50 55 60

Leu Gln Leu Leu Gln Ser Asn Gln Ser Lys Ser Leu Asp Ser Ser Lys
65 70 75 80

His Leu Val Thr Arg Lys Pro Asp Cys Arg Met Glu Thr Cys Phe Asp
85 90 95

Phe Thr Arg Cys Tyr Asp Arg Phe Leu Val Tyr Ile Tyr Pro Pro Glu
100 105 110

Pro Leu Asn Ser Leu Gly Ala Ala Pro Pro Thr Ser Ala Asn Tyr Gln
115 120 125

Lys Ile Leu Thr Ala Ile Gln Glu Ser Arg Tyr Tyr Thr Ser Asp Pro
130 135 140

Thr Ala Ala Cys Leu Phe Val Leu Gly Ile Asp Thr Leu Asp Arg Asp
145 150 155 160

Ser Leu Ser Glu Asp Tyr Val Arg Asn Val Pro Ser Arg Leu Ala Arg
165 170 175

Leu Pro Tyr Trp Asn Asn Gly Arg Asn His Ile Ile Phe Asn Leu Tyr
180 185 190

Ser Gly Thr Trp Pro Asp Tyr Ala Glu Asn Ser Leu Gly Phe Asp Ala
195 200 205

Gly Glu Ala Ile Leu Ala Lys Ala Ser Met Gly Val Leu Gln Leu Arg
210 215 220

His Gly Phe Asp Val Ser Ile Pro Leu Phe His Lys Gln Phe Pro Leu
225 230 235 240

Arg Ala Gly Ala Thr Gly Thr Val Gln Ser Asn Asn Phe Pro Ala Asn
245 250 255

Lys Lys Tyr Leu Leu Ala Phe Lys Gly Lys Arg Tyr Val His Gly Ile
260 265 270

Gly Ser Glu Thr Arg Asn Ser Leu Phe His Leu His Asn Gly Arg Asp
275 280 285

Met Val Leu Val Thr Thr Cys Arg His Gly Lys Ser Trp Arg Glu Leu
290 295 300

Gln Asp Asn Arg Cys Asp Glu Asp Asn Arg Glu Tyr Asp Arg Tyr Asp

305 310 315 320

Tyr Glu Thr Leu Leu Gln Asn Ser Thr Phe Cys Leu Val Pro Arg Gly
325 330 335

Arg Arg Leu Gly Ser Phe Arg Phe Leu Glu Ala Leu Gln Ala Gly Cys
340 345 350

Ile Pro Val Leu Leu Ser Asn Ala Trp Val Leu Pro Phe Glu Ser Lys
355 360 365

Ile Asp Trp Lys Gln Ala Ala Ile Trp Ala Asp Glu Arg Leu Leu Leu
370 375 380

Gln Val Pro Asp Ile Val Arg Ser Ile Pro Ala Glu Arg Ile Phe Ala
385 390 395 400

Leu Arg Gln Gln Thr Gln Val Leu Trp Glu Arg Tyr Phe Gly Ser Ile
405 410 415

Glu Lys Ile Val Phe Thr Thr Phe Glu Ile Ile Arg Glu Arg Leu Pro
420 425 430

Asp Tyr Pro Val Arg Ser Ser Leu Val Trp Asn Ser Ser Pro Gly Ala
435 440 445

Leu Leu Thr Leu Pro Thr Phe Ala Asp Ser Ser Arg Tyr Met Pro Phe
450 455 460

Leu Leu Asn Ser Met Gly Ala Glu Pro Arg His Asn Tyr Thr Ala Val
465 470 475 480

Ile Tyr Val Gln Ile Gly Ala Ala Leu Gly Pro Asn Ala Ala Leu Tyr
485 490 495

Lys Leu Val Arg Thr Ile Thr Lys Ser Gln Phe Val Glu Arg Ile Leu
500 505 510

Val Leu Trp Ala Ala Asp Arg Pro Leu Pro Leu Lys Lys Arg Trp Pro
515 520 525

Pro Thr Ser His Ile Pro Leu His Val Ile Ser Leu Gly Gly Ser Thr
530 535 540

Arg Ser Gln Gly Ala Gly Pro Thr Ser Gln Thr Thr Glu Gly Arg Pro
545 550 555 560

Ser Ile Ser Gln Arg Phe Leu Pro Tyr Asp Glu Ile Gln Thr Asp Ala
565 570 575

Val Leu Ser Leu Asp Glu Asp Ala Ile Leu Asn Thr Asp Glu Leu Asp
580 585 590

Phe Ala Tyr Thr Val Trp Arg Asp Phe Pro Glu Arg Ile Val Gly Tyr
595 600 605

Pro Ala Arg Ala His Phe Trp Asp Asp Ser Lys Asn Ala Trp Gly Tyr
610 615 620

Thr Ser Lys Trp Thr Asn Tyr Tyr Ser Ile Val Leu Thr Gly Ala Ala

```
625                630                635                640

Phe Tyr His Arg Tyr Tyr Asn Tyr Leu Tyr Thr Asn Trp Leu Ser Leu
                645                650                655

Leu Leu Leu Lys Thr Val Gln Gln Ser Ser Asn Cys Glu Asp Ile Leu
            660                665                670

Met Asn Leu Leu Val Ser His Val Thr Arg Lys Pro Pro Ile Lys Val
        675                680                685

Thr Gln Arg Lys Gly Leu Gln Gly Ser Gly Asp Gly Ser Phe Ala Trp
    690                695                700

Asn Asp Pro Asp His Phe Ile Gln Arg Gln Ser Cys Leu Asn Thr Phe
705                710                715                720

Ala Ala Val Phe Gly Tyr Met Pro Leu Ile Arg Ser Asn Leu Arg Met
                725                730                735

Asp Pro Met Leu Tyr Arg Asp Pro Val Ser Asn Leu Arg Lys Lys Tyr
            740                745                750

Arg Gln Ile Glu Leu Val Gly Ser
        755                760
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 1938 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: both
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
  (A) NAME/KEY: CDS
  (B) LOCATION: 82..1797

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GTCTTCCACG TCTGCAGCTC AGCCAGGGCG CGCAGGGCGA GTGGGGTCCA CTGGCGGGTA      60

AAGGGGACCA GGACGGCGAG G ATG GAC GCA CAG ACC TGG CCC GTG GGC TTT       111
                        Met Asp Ala Gln Thr Trp Pro Val Gly Phe
                          1               5                  10

CGC TGC CTC CTC CTT CTG GCC CTG GTT GGG TCC GCC CGC AGC GAG GGC       159
Arg Cys Leu Leu Leu Leu Ala Leu Val Gly Ser Ala Arg Ser Glu Gly
                15                  20                  25

GTG CAG ACC TGC GAA GAA GTT CGG AAA CTT TTC CAG TGG CGG CTG CTG       207
Val Gln Thr Cys Glu Glu Val Arg Lys Leu Phe Gln Trp Arg Leu Leu
            30                  35                  40

GGA GCT GTC AGG GGG CTG CCG GAT TCG CCG CGG GCA GGA CCT GAT CTT       255
```

Gly Ala Val Arg Gly Leu Pro Asp Ser Pro Arg Ala Gly Pro Asp Leu
45 50 55

CAG GTT TGC ATA TCC AAA AAG CCT ACA TGT TGC ACC AGG AAG ATG GAG 303
Gln Val Cys Ile Ser Lys Lys Pro Thr Cys Cys Thr Arg Lys Met Glu
60 65 70

GAG AGA TAT CAG ATT GCG GCT CGC CAG GAT ATG CAG CAG TTT CTT CAA 351
Glu Arg Tyr Gln Ile Ala Ala Arg Gln Asp Met Gln Gln Phe Leu Gln
75 80 85 90

ACG TCC AGC TCT ACA TTA AAG TTT CTA ATA TCT CGA AAT GCG GCT GCT 399
Thr Ser Ser Ser Thr Leu Lys Phe Leu Ile Ser Arg Asn Ala Ala Ala
95 100 105

TTT CAA GAA ACC CTT GAA ACT CTC ATC AAA CAA GCA GAA AAT TAC ACC 447
Phe Gln Glu Thr Leu Glu Thr Leu Ile Lys Gln Ala Glu Asn Tyr Thr
110 115 120

AGT ATA CTT TTT TGC AGT ACC TAC AGG AAC ATG GCC TTG GAG GCT GCT 495
Ser Ile Leu Phe Cys Ser Thr Tyr Arg Asn Met Ala Leu Glu Ala Ala
125 130 135

GCT TCG GTT CAG GAG TTC TTC ACT GAT GTG GGG CTG TAT TTA TTT GGT 543
Ala Ser Val Gln Glu Phe Phe Thr Asp Val Gly Leu Tyr Leu Phe Gly
140 145 150

GCG GAT GTT AAT CCT GAA GAA TTT GTA AAC AGA TTT TTT GAC AGT CTT 591
Ala Asp Val Asn Pro Glu Glu Phe Val Asn Arg Phe Phe Asp Ser Leu
155 160 165 170

TTT CCT CTG GTC TAC AAC CAC CTC ATT AAC CCT GGT GTG ACT GAC AGT 639
Phe Pro Leu Val Tyr Asn His Leu Ile Asn Pro Gly Val Thr Asp Ser
175 180 185

TCC CTG GAA TAC TCA GAA TGC ATC CGG ATG GCT CGC CGG GAT GTG AGT 687
Ser Leu Glu Tyr Ser Glu Cys Ile Arg Met Ala Arg Arg Asp Val Ser
190 195 200

CCA TTT TGT AAT ATT CCC CAA AGA GTA ATG GGA CAG ATG GGG AGG TCC 735
Pro Phe Cys Asn Ile Pro Gln Arg Val Met Gly Gln Met Gly Arg Ser
205 210 215

CTG CTG CCC AGC CGC ACT TTT CTG CAG GCA CTC AAT CTG GGC ATT GAA 783
Leu Leu Pro Ser Arg Thr Phe Leu Gln Ala Leu Asn Leu Gly Ile Glu
220 225 230

GTC ATC AAC ACC ACA GAC TAT CTG CAC TTC TTC AAA GAG TGC AGC AGA 831
Val Ile Asn Thr Thr Asp Tyr Leu His Phe Phe Lys Glu Cys Ser Arg
235 240 245 250

GCC CTC CTG AAG ATG CAA TAC TGC CCG CAC TGC CAA GGC CTG GCG CTC 879
Ala Leu Leu Lys Met Gln Tyr Cys Pro His Cys Gln Gly Leu Ala Leu
255 260 265

ACT AAG CCT TGT ATG GGA TAC TGC CTC AAT GTC ATG CGA GGC TGC CTG 927
Thr Lys Pro Cys Met Gly Tyr Cys Leu Asn Val Met Arg Gly Cys Leu
270 275 280

GCG CAC ATG GCG GAG CTT AAT CCA CAC TGG CAT GCA TAT ATC CGG TCG 975

Ala His Met Ala Glu Leu Asn Pro His Trp His Ala Tyr Ile Arg Ser
285 290 295

TTG GAA GAA CTC TCG GAT GCA ATG CAT GGA ACA TAC GAC ATT GGA CAC 1023
Leu Glu Glu Leu Ser Asp Ala Met His Gly Thr Tyr Asp Ile Gly His
300 305 310

GTG CTG CTG AAC TTT CAC TTG CTT GTT AAT GAT GCT GTG TTA CAG GCT 1071
Val Leu Leu Asn Phe His Leu Leu Val Asn Asp Ala Val Leu Gln Ala
315 320 325 330

CAC CTC AAT GGA CAA AAA TTA TTG GAA CAG GTA AAT AGG ATT TGT GGC 1119
His Leu Asn Gly Gln Lys Leu Leu Glu Gln Val Asn Arg Ile Cys Gly
335 340 345

CGC CCT GTA AGA ACA CCC ACA CAA AGC CCC CGT TGT TCT TTT GAT CAG 1167
Arg Pro Val Arg Thr Pro Thr Gln Ser Pro Arg Cys Ser Phe Asp Gln
350 355 360

AGC AAA GAG AAG CAT GGA ATG AAG ACC ACC ACA AGG AAC AGT GAA GAG 1215
Ser Lys Glu Lys His Gly Met Lys Thr Thr Thr Arg Asn Ser Glu Glu
365 370 375

ACG CTT GCC AAC AGA AGA AAA GAA TTT ATC AAC AGC CTT CGA CTG TAC 1263
Thr Leu Ala Asn Arg Arg Lys Glu Phe Ile Asn Ser Leu Arg Leu Tyr
380 385 390

AGG TCA TTC TAT GGA GGT CTA GCT GAT CAG CTT TGT GCT AAT GAA TTA 1311
Arg Ser Phe Tyr Gly Gly Leu Ala Asp Gln Leu Cys Ala Asn Glu Leu
395 400 405 410

GCT GCT GCA GAT GGA CTT CCC TGC TGG AAT GGA GAA GAT ATA GTA AAA 1359
Ala Ala Ala Asp Gly Leu Pro Cys Trp Asn Gly Glu Asp Ile Val Lys
415 420 425

AGT TAT ACT CAG CGT GTG GTT GGA AAT GGA ATC AAA GCC CAG TCT GGA 1407
Ser Tyr Thr Gln Arg Val Val Gly Asn Gly Ile Lys Ala Gln Ser Gly
430 435 440

AAT CCT GAA GTC AAA GTC AAA GGA ATT GAT CCT GTG ATA AAT CAG ATT 1455
Asn Pro Glu Val Lys Val Lys Gly Ile Asp Pro Val Ile Asn Gln Ile
445 450 455

ATT GAT AAA CTG AAG CAT GTT GTT CAG TTG TTA CAG GGT AGA TCA CCC 1503
Ile Asp Lys Leu Lys His Val Val Gln Leu Leu Gln Gly Arg Ser Pro
460 465 470

AAA CCT GAC AAG TGG GAA CTT CTT CAG CTG GGC AGT GGT GGA GGC ATG 1551
Lys Pro Asp Lys Trp Glu Leu Leu Gln Leu Gly Ser Gly Gly Gly Met
475 480 485 490

GTT GAA CAA GTC AGT GGG GAC TGT GAT GAT GAA GAT GGT TGC GGG GGA 1599
Val Glu Gln Val Ser Gly Asp Cys Asp Asp Glu Asp Gly Cys Gly Gly
495 500 505

TCA GGA AGT GGA GAA GTC AAG AGG ACA CTG AAG ATC ACA GAC TGG ATG 1647
Ser Gly Ser Gly Glu Val Lys Arg Thr Leu Lys Ile Thr Asp Trp Met
510 515 520

CCA GAT GAT ATG AAC TTC AGT GAT GTA AAG CAA ATC CAT CAA ACA GAC 1695

Pro Asp Asp Met Asn Phe Ser Asp Val Lys Gln Ile His Gln Thr Asp
525 530 535

ACT GGC AGT ACT TTA GAC ACA ACA GGA GCA GGA TGT GCA GTG GCG ACT 1743
Thr Gly Ser Thr Leu Asp Thr Thr Gly Ala Gly Cys Ala Val Ala Thr
540 545 550

GAA TCT ATG ACA TTC ACT CTG ATA AGT GTG GTG ATG TTA CTT CCC GGG 1791
Glu Ser Met Thr Phe Thr Leu Ile Ser Val Val Met Leu Leu Pro Gly
555 560 565 570

ATT TGG TAACTGAACT CTTCTGTCCT GACATACCTT ACTGAAGTCT CGATTTCTTC 1847
Ile Trp

TCTCTCTGCA TATGCCTGGA ATAAGAGATC CTTTTTCAAT GTAACAATTA TATTTATGAA 1907

AAGATATGTT ACACTAACTT CCAGAAGCCA A 1938

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 572 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Asp Ala Gln Thr Trp Pro Val Gly Phe Arg Cys Leu Leu Leu Leu
1 5 10 15

Ala Leu Val Gly Ser Ala Arg Ser Glu Gly Val Gln Thr Cys Glu Glu
20 25 30

Val Arg Lys Leu Phe Gln Trp Arg Leu Leu Gly Ala Val Arg Gly Leu
35 40 45

Pro Asp Ser Pro Arg Ala Gly Pro Asp Leu Gln Val Cys Ile Ser Lys
50 55 60

Lys Pro Thr Cys Cys Thr Arg Lys Met Glu Glu Arg Tyr Gln Ile Ala
65 70 75 80

Ala Arg Gln Asp Met Gln Gln Phe Leu Gln Thr Ser Ser Ser Thr Leu
85 90 95

Lys Phe Leu Ile Ser Arg Asn Ala Ala Ala Phe Gln Glu Thr Leu Glu
100 105 110

Thr Leu Ile Lys Gln Ala Glu Asn Tyr Thr Ser Ile Leu Phe Cys Ser
115 120 125

Thr Tyr Arg Asn Met Ala Leu Glu Ala Ala Ala Ser Val Gln Glu Phe
130 135 140

Phe Thr Asp Val Gly Leu Tyr Leu Phe Gly Ala Asp Val Asn Pro Glu
145 150 155 160

Glu Phe Val Asn Arg Phe Phe Asp Ser Leu Phe Pro Leu Val Tyr Asn
165 170 175

His Leu Ile Asn Pro Gly Val Thr Asp Ser Ser Leu Glu Tyr Ser Glu
180 185 190

Cys Ile Arg Met Ala Arg Arg Asp Val Ser Pro Phe Cys Asn Ile Pro
195 200 205

Gln Arg Val Met Gly Gln Met Gly Arg Ser Leu Leu Pro Ser Arg Thr
210 215 220

Phe Leu Gln Ala Leu Asn Leu Gly Ile Glu Val Ile Asn Thr Thr Asp
225 230 235 240

Tyr Leu His Phe Phe Lys Glu Cys Ser Arg Ala Leu Leu Lys Met Gln
245 250 255

Tyr Cys Pro His Cys Gln Gly Leu Ala Leu Thr Lys Pro Cys Met Gly
260 265 270

Tyr Cys Leu Asn Val Met Arg Gly Cys Leu Ala His Met Ala Glu Leu
275 280 285

Asn Pro His Trp His Ala Tyr Ile Arg Ser Leu Glu Glu Leu Ser Asp
290 295 300

Ala Met His Gly Thr Tyr Asp Ile Gly His Val Leu Leu Asn Phe His
305 310 315 320

Leu Leu Val Asn Asp Ala Val Leu Gln Ala His Leu Asn Gly Gln Lys
325 330 335

Leu Leu Glu Gln Val Asn Arg Ile Cys Gly Arg Pro Val Arg Thr Pro
340 345 350

Thr Gln Ser Pro Arg Cys Ser Phe Asp Gln Ser Lys Glu Lys His Gly
355 360 365

Met Lys Thr Thr Thr Arg Asn Ser Glu Glu Thr Leu Ala Asn Arg Arg
370 375 380

Lys Glu Phe Ile Asn Ser Leu Arg Leu Tyr Arg Ser Phe Tyr Gly Gly
385 390 395 400

Leu Ala Asp Gln Leu Cys Ala Asn Glu Leu Ala Ala Ala Asp Gly Leu
405 410 415

Pro Cys Trp Asn Gly Glu Asp Ile Val Lys Ser Tyr Thr Gln Arg Val
420 425 430

Val Gly Asn Gly Ile Lys Ala Gln Ser Gly Asn Pro Glu Val Lys Val
435 440 445

Lys Gly Ile Asp Pro Val Ile Asn Gln Ile Ile Asp Lys Leu Lys His
450 455 460

Val Val Gln Leu Leu Gln Gly Arg Ser Pro Lys Pro Asp Lys Trp Glu
465 470 475 480

Leu Leu Gln Leu Gly Ser Gly Gly Gly Met Val Glu Gln Val Ser Gly
                485   490   495

Asp Cys Asp Asp Glu Asp Gly Cys Gly Gly Ser Gly Ser Gly Glu Val
            500   505   510

Lys Arg Thr Leu Lys Ile Thr Asp Trp Met Pro Asp Asp Met Asn Phe
        515   520   525

Ser Asp Val Lys Gln Ile His Gln Thr Asp Thr Gly Ser Thr Leu Asp
    530   535   540

Thr Thr Gly Ala Gly Cys Ala Val Ala Thr Glu Ser Met Thr Phe Thr
545   550   555   560

Leu Ile Ser Val Val Met Leu Leu Pro Gly Ile Trp
                565   570

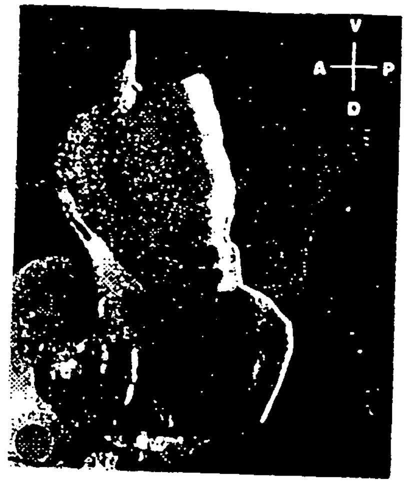


Fig. 1A

Fig. 1B

Fig. 1C

Fig. 1D

Fig. 1E

Fig. 1F

Fig. 1G

Fig. 1H

Fig. 1I

Fig. 1J

Fig. 1K

Fig. 2A

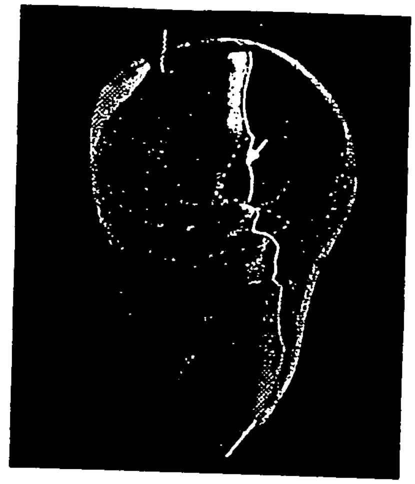


Fig. 2B

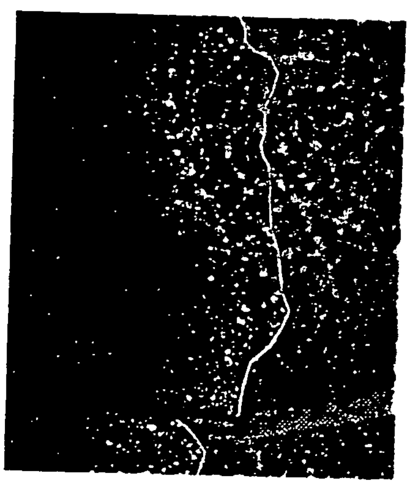


Fig. 2C

Fig. 2D

Fig. 2E

Fig. 2F

Fig. 2G

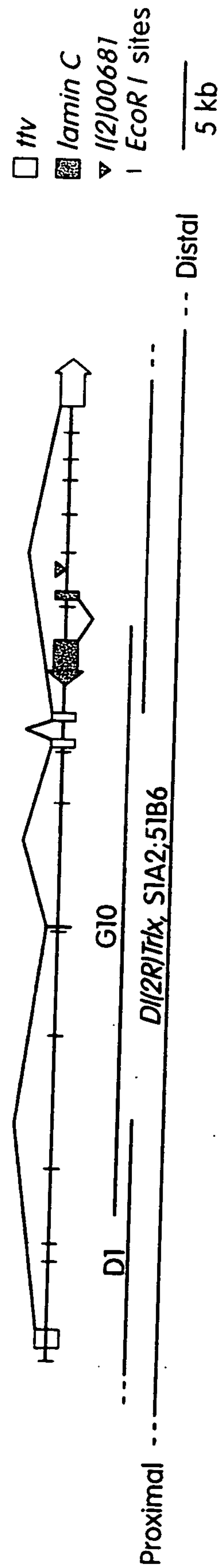

ttv
lamin C
l(2)00681
EcoR I sites
5 kb
Distal
G10
D1
Df(2R)Trlx, 51A2;51B6
Proximal

Fig. 3A

```
MQAKKRY LVFV..SCAFLAYAYFGGYRLKVSPLRPRRAQHESAKDGGVQP.........HEQLPSFLGAHDMQELQLLQSNQSKSLDSSKHLVTRKPDC Ttv
MQAKKRYFILLSAGSCLALLF.YFGGLQFRASRSHSRREEH.SGRNGLHHPSPDHFWPRFPEPLRPFVPWDQLENEDSSVHISPRQKRDANSSIYKGKKC Ext-1

RMETCFDFTRCYDR.FLVY YPPEPLNSLGAAPPTSANYQKILTAIQESRYYTSDPTAACLFVLGIDTLDRDSLSEDYVRNVPSRLARLPYWNNGRNHII Ttv
RMESCFDFTLCKKNGFKVYVYPQQKGEKIA......ESYQNILAAIEGSRFYTSDPSQACLFVLSLDTLDRDQLSPQYVHNLRSKVQSLHLWNNGRNHLI Ext-1

FNLYSGTWPDYAENSLGFDAGEA LAKASMGVLQLRHGFDVSIPLFHKQFPLRAGATGTVQSNNFPANKKYLLAFKGKRYVHGIGSETRNSLFHLHNGRD Ttv
FNLYSGTWPDYTED-VGFD GQAMLAKAS STENFRPNFDVSIPLFSKDHPRTGGERGFLKFNTIPPLRKYMLVFKGKRYLTGIGSDTRNALYHVHNGED Ext-1

MVLVTTCRHGKSWRELQDNRCDEDNREYDRYDYETLLQNSTFCLVPRGRRLGSFRFLEALQAGCIPVLLSNAWVLPFESKIDWKQAA WADERLLLQVPD Ttv
VVLLTTCKHGKDWQKHKDSRCDRDNTEYEKYDYREMLHNATFCLVPRGRRLGSFRFLEALQAACVPVMLSNGWELPFSEVINWNQAAVIGDERLLLQ PS Ext-1

IVRSIPAERIFALRQQTQVLWERYFGS EKIVFTTFEIIRERLPDYPVRSSLVWNSSPGALLTLPTFADSSRYMPFLLNSMGAEPRHNYTAVIYVQIGAA Ttv
TIRSIHQDKILALRQQTQFLWEAYFSSVEKIVLTTLEIIQDR FKHISRNSL WNKHPGGLFVLPQYSSYLGDFPYYYANLGLKPPSKFTAVIHA.VTPL Ext-1

LGPNAALYKLVRTITKSQFVERILVLWAADRPLPLKKRWPPTSHIPLHV SLGGSTRSQGAGPTSQTTEGRPSISQRFLPYDEIQTDAVLSLDEDAILNT Ttv
VSQSQPVLKLLVAAAKSQYCAQI VLWNCDKPLPAKHRWPATA...VPVVVIEGESKV...............MSSRFLPYDNI TDAVLSLDEDTVLST Ext-1

DELDFAYTVWRDFPERIVGYPARAHFWDDSKNAWGYTSKWTNYYS VLTGAAFYHRYYNYLYTNWLSLLLLKTVQQSSNCEDILMNLLVSHVTRKPPIKV Ttv
TEVDFAFTVWQSFPERIVGYPARSHFWDNSKERWGYTSKWTNDYSMVLTGAA YHKYYHYLYSHYLPASLKNMVOQLANCEDILMNFLVSAVTKLPPIKV Ext-1

TQRK....GLQGSGDGSFAWNDPDHF QRQSCLNTFAAVFGYMPLIRSNLRMDPMLYRDPVSNLRKKYRQIELVGS Ttv
TQKKQYKETMMGQTSRASRWADPDHFAQRQSCMNTFASWFGYMPLIHSQMRLDPVLFKDQVS LRKKYRQIERL Ext-1
```

Fig. 3B

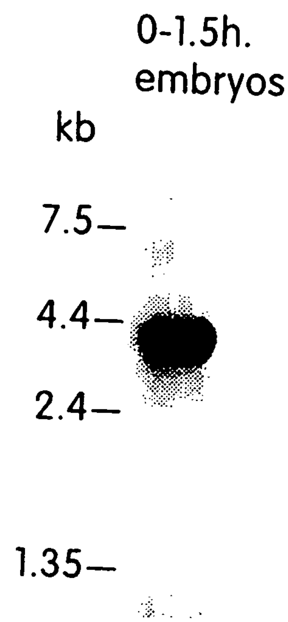


Fig. 3C

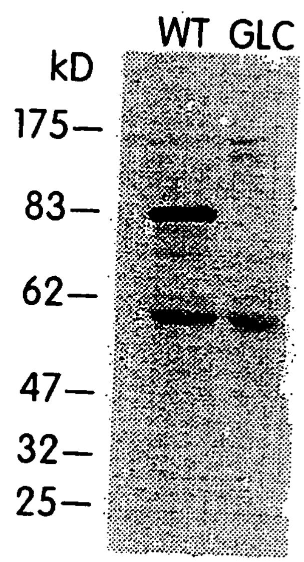


Fig. 3D

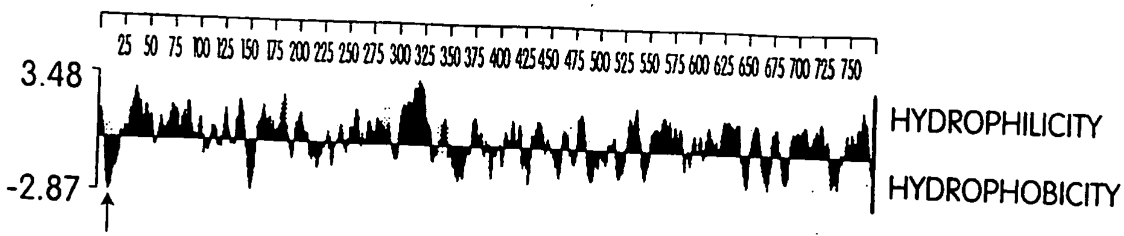


Fig. 4A

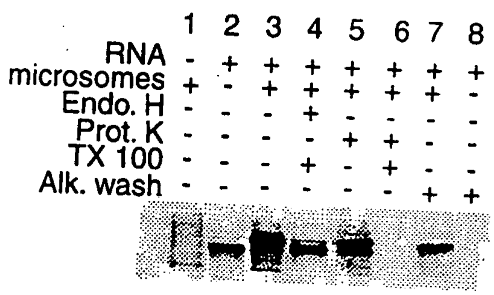


Fig. 4B

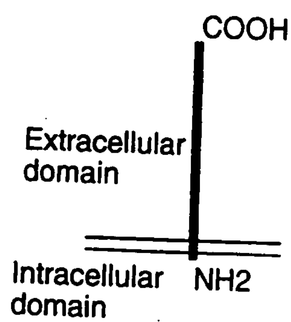


Fig. 4C

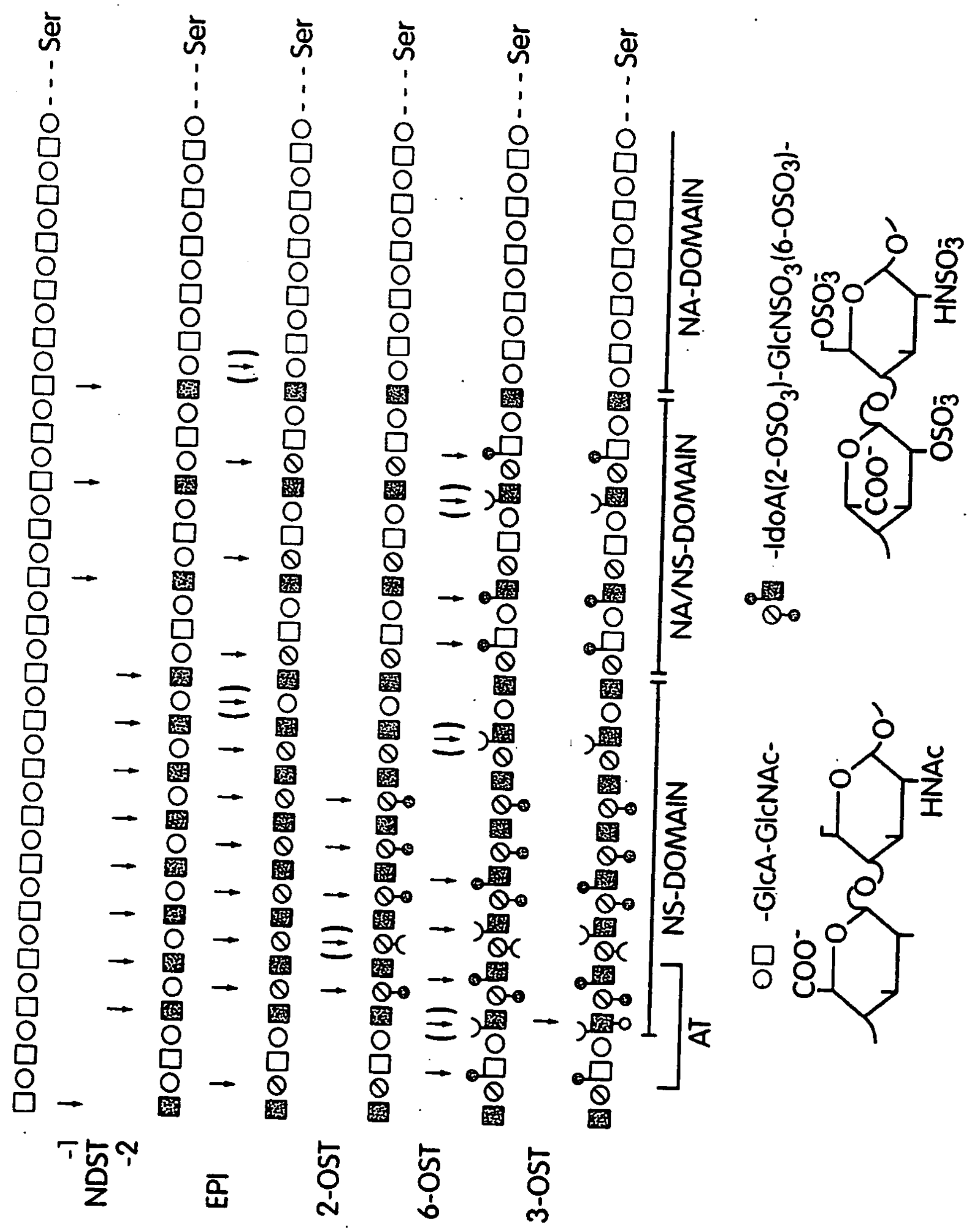

-1
NDST
-2
EPI
2-OST
6-OST
3-OST
Ser
AT
NS-DOMAIN
NA/NS-DOMAIN
NA-DOMAIN
-GlcA-GlcNAc-
-IdoA(2-OSO3)-GlcNSO3(6-OSO3)-
COO-
HNAc
OSO3-
HNSO3

Fig. 5

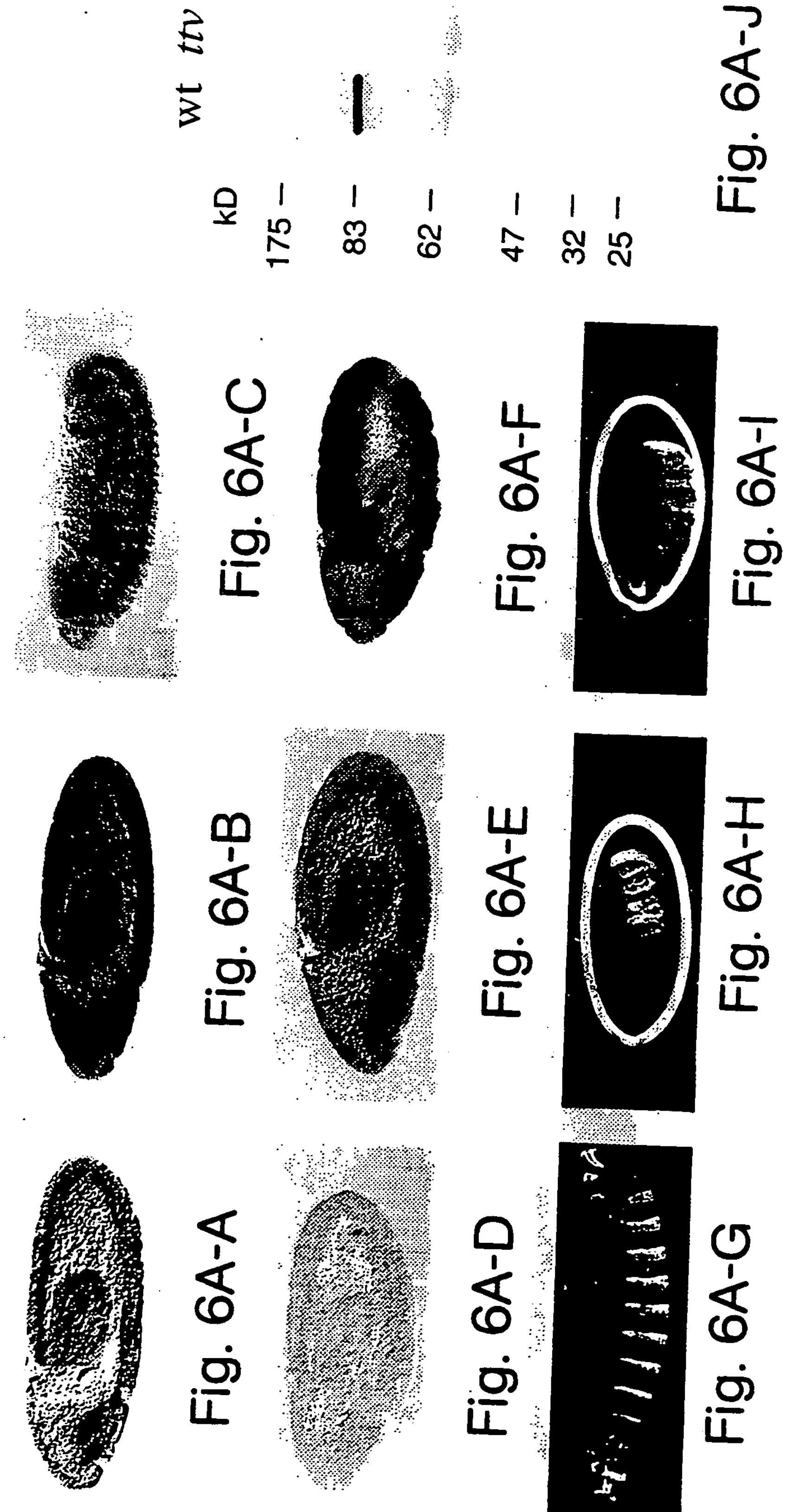


Fig. 6A-A

Fig. 6A-B

Fig. 6A-C

Fig. 6A-D

Fig. 6A-E

Fig. 6A-F

Fig. 6A-G

Fig. 6A-H

Fig. 6A-I

Fig. 6A-J

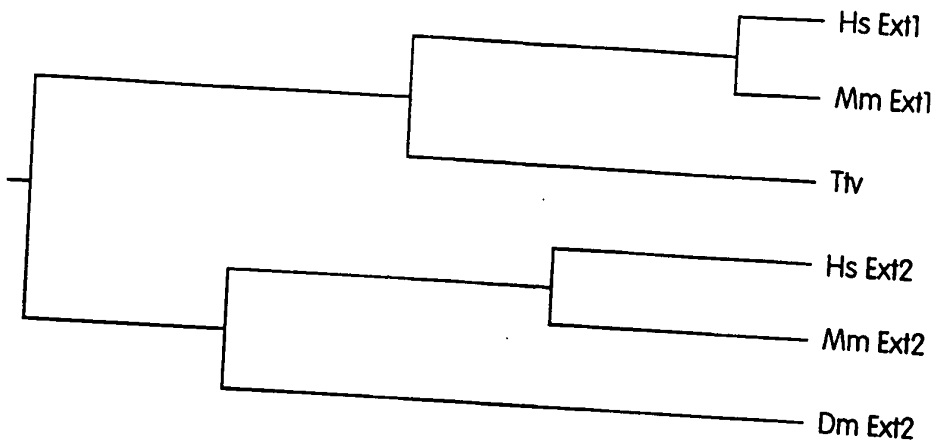

Hs Ext1
Mm Ext1
Ttv
Hs Ext2
Mm Ext2
Dm Ext2

Fig. 6B

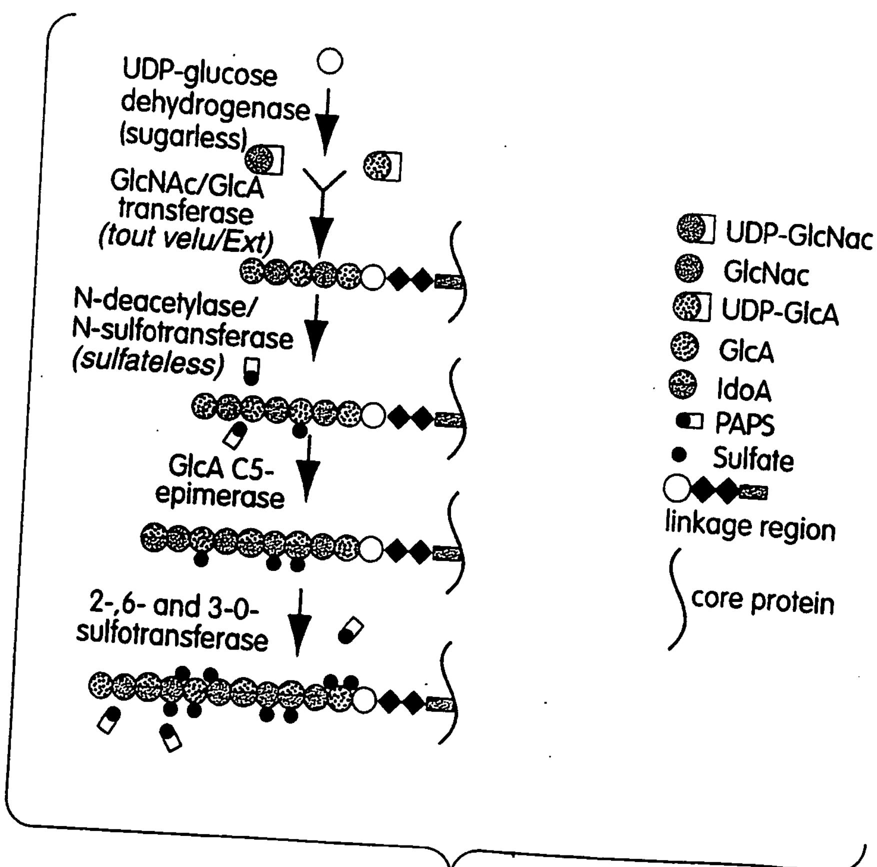

UDP-glucose dehydrogenase (sugarless)
GlcNAc/GlcA transferase (tout velu/Ext)
N-deacetylase/ N-sulfotransferase (sulfateless)
GlcA C5- epimerase
2-,6- and 3-0- sulfotransferase
UDP-GlcNac
GlcNac
UDP-GlcA
GlcA
IdoA
PAPS
Sulfate
linkage region
core protein

Fig. 6C

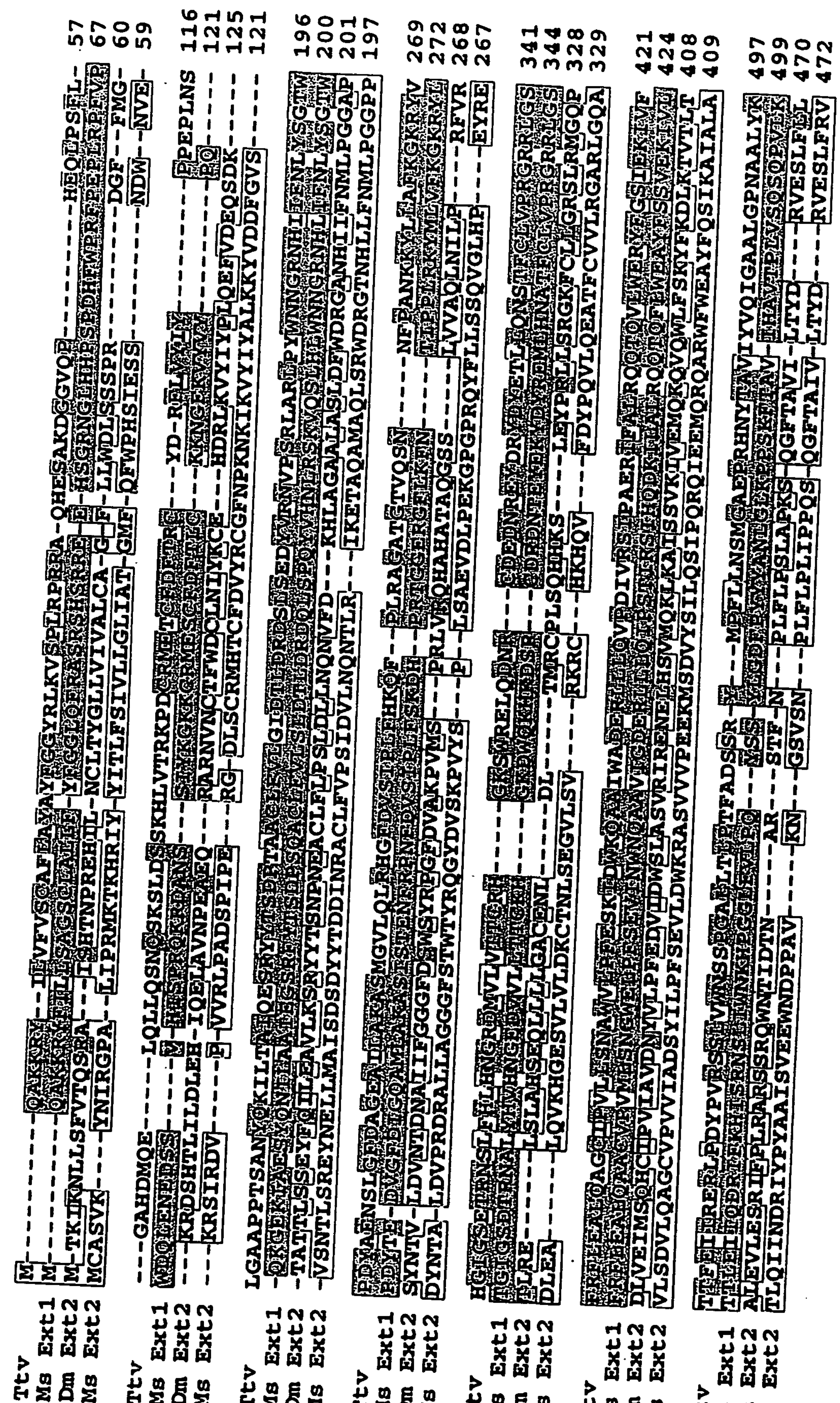


Fig. 7

```
Ttv      EVRTITKSQFVERTLVLWA-ADRPPPLKKRWPPTSHIPLHVISLGGSTRSQGAGPTSQTTEGRPSISQRFLPYDEIQTDA 576
Ms Ext1  LLVAAAKSQYCAQLLVLWN-SDKPLPAKIRWPAT---AVPVVVEEGES---------------KVMSSRFLPYDNIITDA 560
Dm Ext2  IQKLAVVPSLQSILVIWNNQKKSPPHLSTFPSIS---KPLKIRQTKEN---------------KLSNRFYPYPEIETEAI 532
Ms Ext2  ITEVSKVPSLSKLLVVWNNQNKNPPEDSLWPKIR---VPLKVVRTAEN---------------KLSNRFFPYDEIETEAV 534

Ttv      VLSLDEDA-IINTDELDFAYTVWRDEPERIVGYPARAHFWDDSKNAWGYTSKWTNYYSIVLTGAAFYHRYYNYLYTNWLS 655
Ms Ext1  VLSLDEDT-VLSTTEVDFAFTVWQSEPERIVGYPARSHFWDNSKERWGYTSKWTNDYSMVLTGAAIYHKYYHYLYSHYLP 639
Dm Ext2  LTIDDDIIMLTTDELDFGYEVWREFPDHIVGFPSRIHVWENVTMRWHYESEWTNQISMVLTGAAFHHKYWSHMYTHAMPG 612
Ms Ext2  LAIDDDIIMLTSDELQFGYEVWREFPDRLVGYPGRLHLWDHEMNKWKYESEWTNEVSMVLTGAAFYNKYFNYLYTYKMPG 614

Ttv      LLELKTVQQSSNCEDILMNLEVSHVTRKPPIKVTQR----KGLQGSGDGSFAWNDPDHFIQRQSCLNTFAAVFGYMPLIR 731
Ms Ext1  ASLKNMVDQLANCEDILMNFLVSAVTKIPPIKVTQKKQYKETMMGQTSRASRWADPDHEAQRQSCMNTEASWFGYMPLIH 719
Dm Ext2  DIKDWVDEHMNCEDIAMNFLVANITNNPPIKVTPRKKFKCPECTNTEMLSA--DLNHMRERSACIDRFSKIYGRMPLRTV 690
Ms Ext2  DIKNWVDAHMNCEDIAMNFLVANVTGKAVIKVTPRKKFKCPECTAIDGLSL--DQTHMVERSECINKFASVFGTMPLKVV 692

Ttv      SNLRMDRMLYRDPVSNLRKKYRQTELVGS-.
Ms Ext1  SQMRLDPVLFKDQVSILRKKYRDI---ER-[illegible]
Dm Ext2  EFRADPVLFRDNFPDKLKRYNDIG---SL.
Ms Ext2  EHRADPVLYKDDFPEKLKSFPNIG---SL
```

▩ AMINO ACID RESIDUE IDENTICAL TO Hs Ext1

□ AMINO ACID RESIDUE IDENTICAL TO Hs Ext2

Fig. 7 cont.

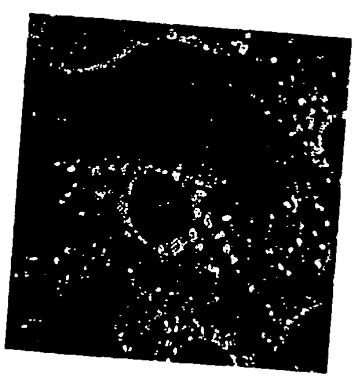


Fig. 8A

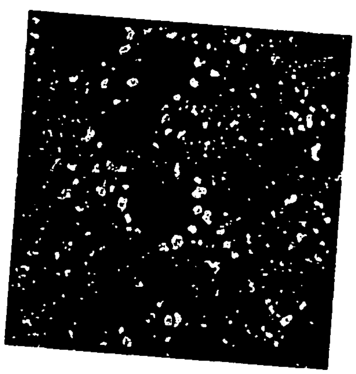


Fig. 8B

Fig. 8C

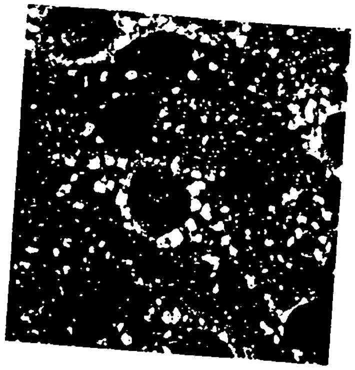


Fig. 8D

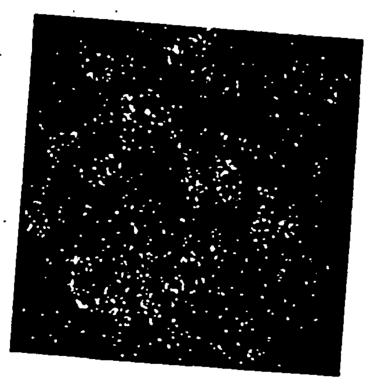


Fig. 8E

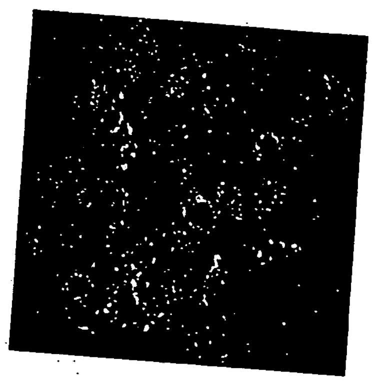


Fig. 8F

Fig. 8G

Fig. 8H

Fig. 9A

Fig. 9B

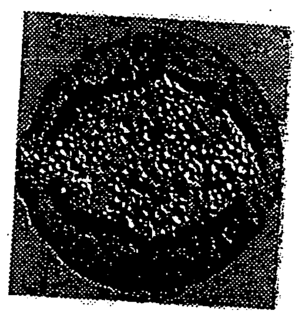


Fig. 9C

Fig. 9D

Fig. 10A

Fig. 10B

Fig. 10C

Fig. 10D

Fig. 10E

Fig. 10F

Fig. 10G

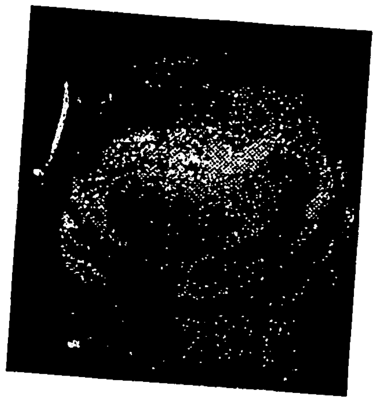


Fig. 10H

Fig. 10I

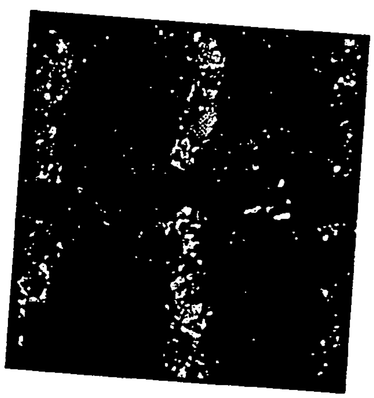


Fig. 10J

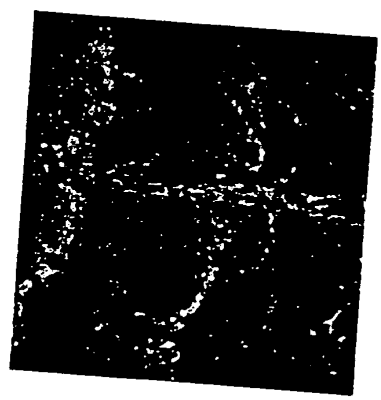


Fig. 10K

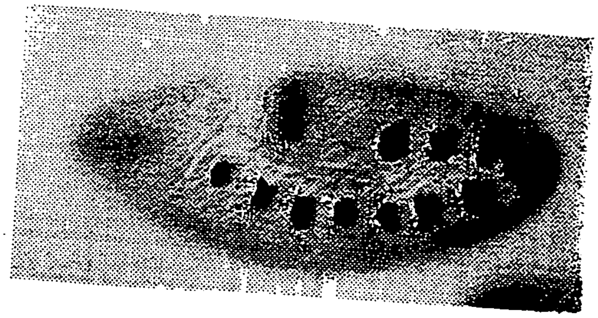


Fig. 11A

Fig. 11B

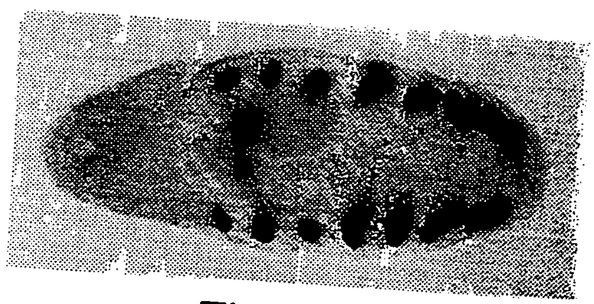


Fig. 11C

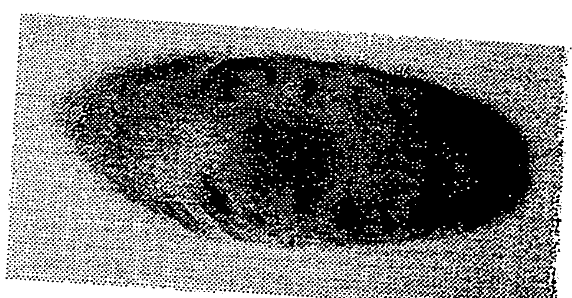


Fig. 11D

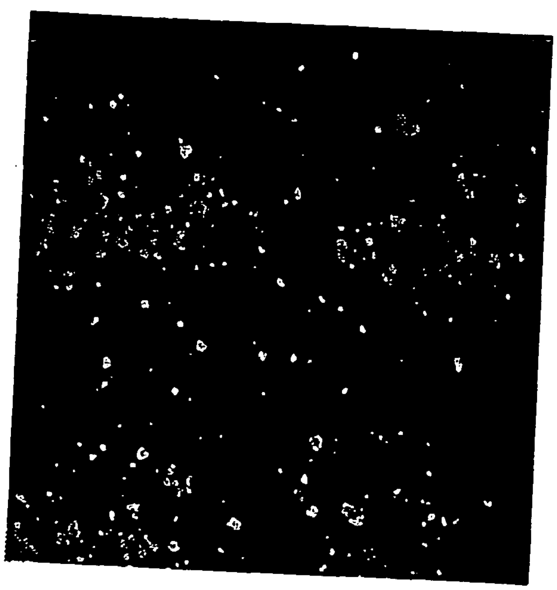


Fig. 11E

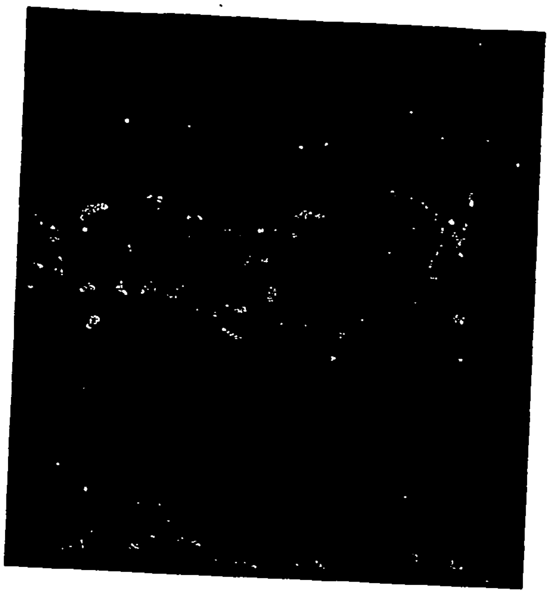


Fig. 11F

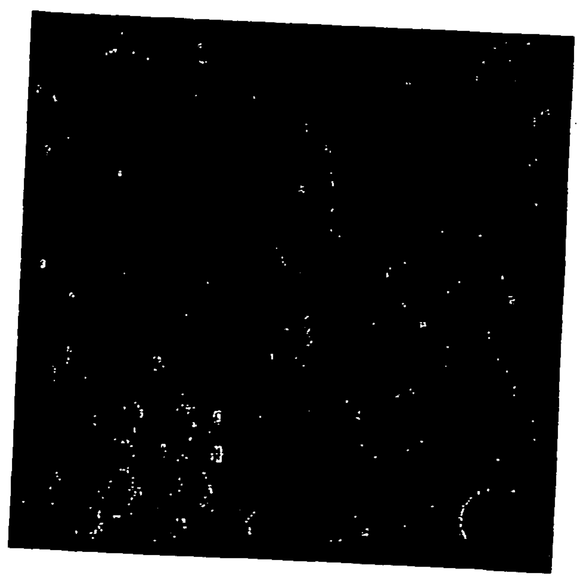


Fig. 11G

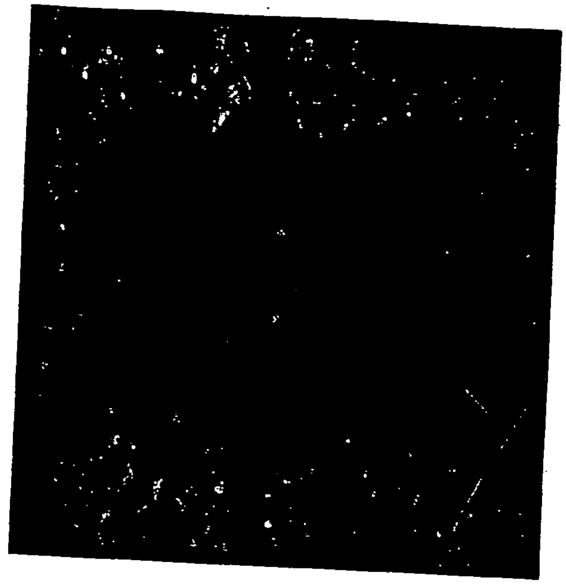


Fig. 11H

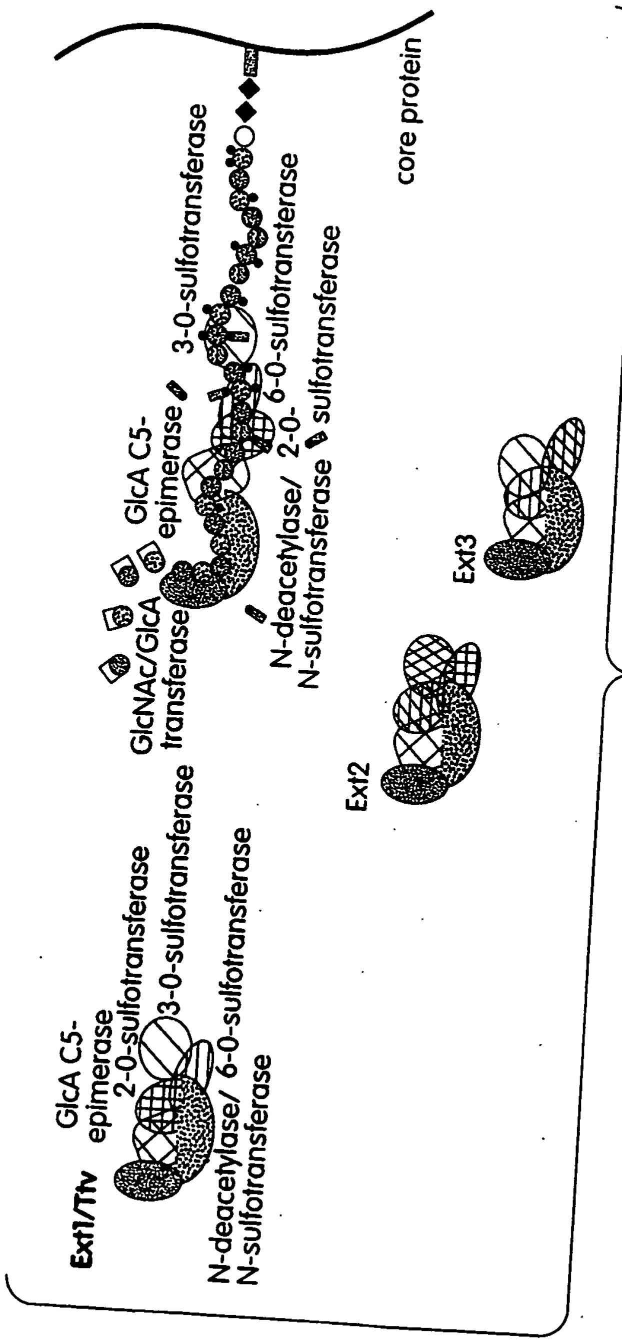

Ext1/Ttv
GlcA C5-
epimerase
2-0-sulfotransferase
3-0-sulfotransferase
N-deacetylase/ 6-0-sulfotransferase
N-sulfotransferase
GlcNAc/GlcA
transferase
GlcA C5-
epimerase
3-0-sulfotransferase
6-0-sulfotransferase
N-deacetylase/ 2-0-
N-sulfotransferase sulfotransferase
core protein
Ext2
Ext3

Fig. 12A

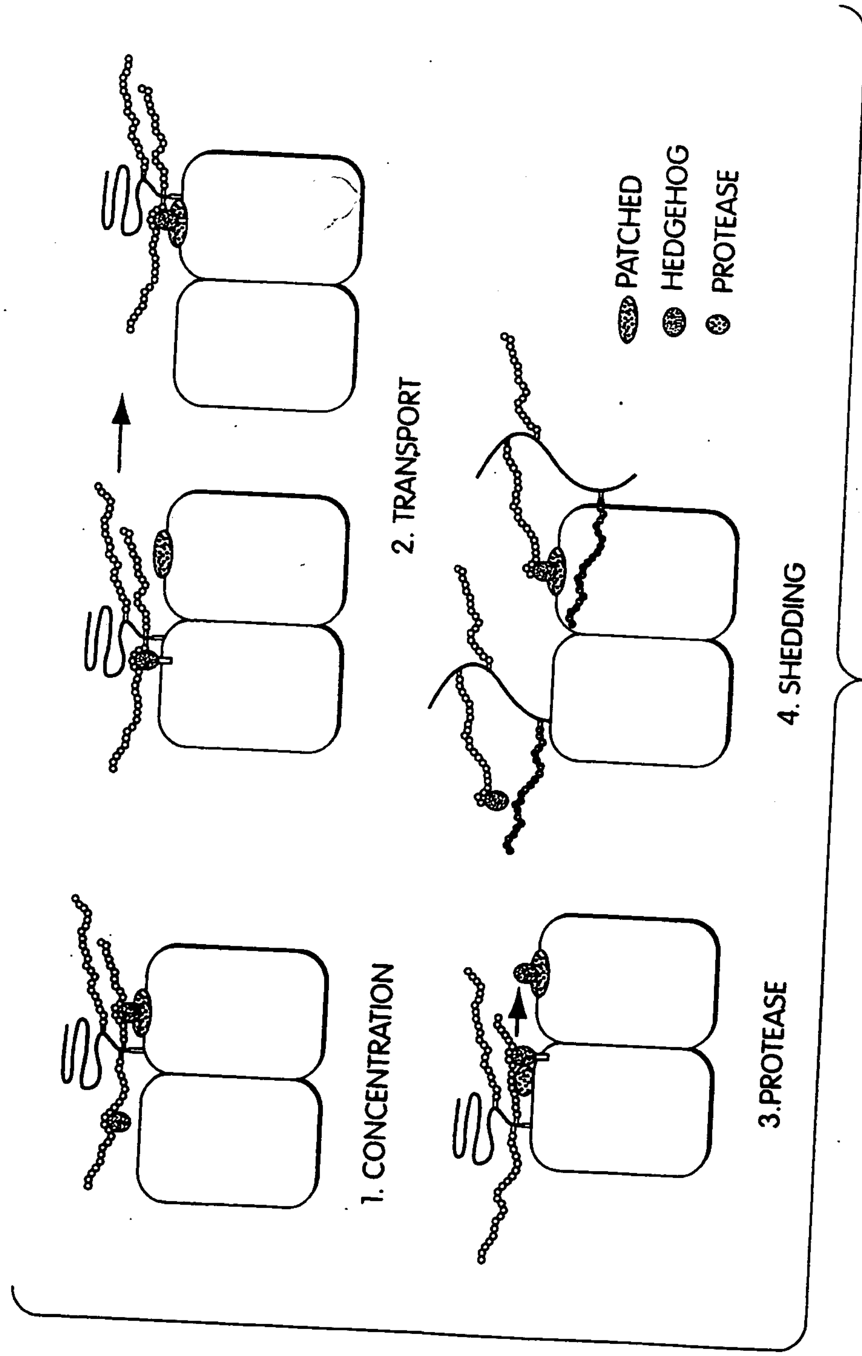

1. CONCENTRATION
2. TRANSPORT
3.PROTEASE
4. SHEDDING
PATCHED
HEDGEHOG
PROTEASE

Fig. 12B